

Evaluating Multimodal Fusion Strategies for Resilient Agricultural Sensing Systems

A PROJECT REPORT

Submitted by

ABHAY SHAJI VALIYAPARAMBIL [RA2112704010006]

PONNURI ANIRUDDHA [RA2112704010015]

Under the Guidance of

Dr. K Sornalakshmi

Associate Professor, Department Data Science and Business System

in partial fulfillment of the requirements for the degree of

MASTER OF TECHNOLOGY [Integrated]

in

COMPUTER SCIENCE AND ENGINEERING

with specialization in Data Science



**DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS
COLLEGE OF ENGINEERING AND TECHNOLOGY SRM
INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR- 603 203**

MAY 2025



Department of Data Science and Business System
SRM Institute of Science & Technology
Own Work* Declaration Form

Degree/ Course : MTech[Integrated]
Student Name : Abhay Shaji Valiyaparambil, Aniruddha Ponnuri
Registration Number : RA2112704010006, RA2112704010015

Title of Work : Evaluating Multimodal Fusion Strategies for Resilient Agricultural Sensing Systems

We hereby certify that this assessment complies with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is our own except where indicated, and that We have met the following conditions:

- Clearly referenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc.)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that we have received from others (e.g. Fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

We understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:

We are aware of and understand the University's policy on Academic misconduct and plagiarism and we certify that this assessment is our own work, except where indicated by referring, and that I have followed the good academic practices noted above.

Abhay Shaji Valiyaparambil
[RA2112704010006]

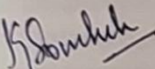
Ponnuri Aniruddha
[RA2112704010015]



**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603 203**

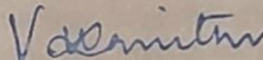
BONAFIDE CERTIFICATE

Certified that 21CSP401L - Project report titled "Evaluating Multimodal Fusion Strategies for Resilient Agricultural Sensing Systems" is the bonafide work of "ABHAY SHAJI VALIYAPARAMBIL [RA2112704010006], ANIRUDDHA PONNURI [RA2112704010006]" who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.


SIGNATURE

Dr.K Sornalakshmi
SUPERVISOR
Associate Professor
Department of Data Science And
Business Systems

EXAMINER 1


SIGNATURE

Dr. Kavitha V
PROFESSOR & HEAD
DEPARTMENT OF
Data Science and Business Systems

EXAMINER 2

ACKNOWLEDGEMENTS

We express our humble gratitude to Dr. C. Muthamizhchelvan, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. Leenus Jesu Martin M**, Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We encompass our sincere thanks to, **Dr. C. Lakshmi**, Professor and Associate Chairperson, School of Computing, SRM Institute of Science and Technology, for her invaluable support.

We are incredibly grateful to our Head of the Department, **Dr. KAVITHA V**, Professor and Head, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators, Panel Head, and Panel Members Department of Data Science and Business Systems, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor **Dr. K ShanthaKumari**, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. K. Sornalakshmi**, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under her mentorship. She provided us with freedom and support to explore the research topics of our interest. Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff members of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement

Aniruddha Ponnuri [RA2112703010018]

Abhay Shaji Valiyaparambil [RA2112703010020]

ABSTRACT

The integration of multimodal streams of data in agriculture, specifically time series sensor data and crop images, is an opportunity and a challenge for predictive analytics in precision agriculture. In this paper, we explore systematically novel data fusion methods to improve molecular property prediction and crop health analysis by integrating heterogeneous modalities. We utilize an uncommon dataset that consists of synchronized measurements from the soil sensor (moisture, pH, NPK nutrients) and high-resolution images of the fields, which are obtained over different temporal and environmental conditions. We compare three state-of-the-art fusion networks: Multimodal Data Fusion-based Graph Contrastive Learning (MDFCL), the Graph-Structured & Interlaced-Masked Fusion Network (GSIFN), and Perceiver IO, a latent-bottleneck multimodal transformer. Each model encapsulates a different philosophy of multimodal representation. MDFCL constructs individual graphs for every modality and aligns their representations with contrastive objectives with no explicit supervision in order to encourage robustness. GSIFN weaves modality-specific masks into a unified transformer, maintaining higher-order cross-modal interactions but without parameter duplication. Perceiver IO, by contrast, employs an asymmetric attention bottleneck to compress heterogeneous inputs into a dense latent space, providing scalability and flexibility with arbitrary combinations of modalities. We compare these architectures on their predictive accuracy for key agronomic traits, e.g., potassium content in the soil, under standard and degraded conditions. We evaluate performance using a broad suite of metrics: accuracy, cross-entropy loss, corruption robustness to synthetic corruption (sensor dropout, image noise, modality ablation). We also investigate model complexity, inference latency, and computational expense to inform real-world deployment. Empirical findings verify that all three fusion approaches outperform unimodal baselines but differ substantially from one another in terms of robustness with corrupt or missing data. Although MDFCL is more stable under sensor corruption, GSIFN is more robust under partial image loss. Perceiver IO is competitively accurate with the advantage of quasi-linear complexity with respect to input size and hence appropriate for large-scale, real-time applications. Cross-validation is employed to verify robustness of such findings under diverse field conditions. This work not only compares state-of-the-art multimodal fusion techniques in a novel agriculture context but also provides practical advice to practice experts who need to build robust, scalable solutions for sensor and image fusion. By enforcing rigorous comparison of architectural paradigms and quantification of their trade-offs, we pave the way for future advances in data-driven agriculture. The techniques and results described here are a template for applying multimodal fusion to other domains with complex, heterogeneous streams of data.

TABLE OF CONTENTS

ABSTRACT		v
TABLE OF CONTENTS		vi
LIST OF FIGURES		viii
ABBREVIATIONS		ix
CHAPTER NO.	TITLE	PAGE NO.
1	INTRODUCTION	10
	1.1 Introduction to Project	10
	1.2 Problem Statement	10
	1.3 Motivation	11
	1.4 Sustainable Development Goal of the Project	11
2	LITERATURE SURVEY	13
	2.1 Overview of the Research Area	13
	2.2 Existing Models and Frameworks	14
	2.3 Limitations Identified from Literature Survey (Research Gaps)	15
	2.4 Research Objectives	17
	2.5 Product Backlog (Key user stories with Desired outcomes)	18
	2.5 Plan of Action (Project Road Map)	19
3	SPRINT PLANNING AND EXECTION METHODOLOGY	22
	3.1 SPRINT I	22
	3.1.1 Objectives with user stories of Sprint I	11
	3.1.2 Functional Document	12
	3.1.3 Architecture Document	13
	3.1.4 Outcome of objectives/ Result Analysis	15
	3.1.5 Sprint Retrospective	15
	3.2 SPRINT II	25
	3.2.1 Objectives with user stories of Sprint II	16
	3.2.2 Functional Document	17

3.2.3 Architecture Document	18
3.2.4 Outcome of objectives/ Result Analysis	20
3.2.5 Sprint Retrospective	21
4 PROPOSED SYSTEM	29
4.1 Proposed Solution	29
4.2 System Architecture	30
5 RESULTS AND DISCUSSIONS	33
5.1 Performance Overview	33
5.2 Comparative Analysis	33
5.3 Resource Efficiency	33
5.4 Summary of Findings	33
6 CONCLUSION AND FUTURE ENHANCEMENT	35
6.1 Conclusion	35
6.2 Future Enhancements	35
REFERENCES	36
APPENDIX	37
A CODING	37
B CONFERENCE PUBLICATION	45
C PLAGIARISM REPORT	46

LIST OF FIGURES

CHAPTER NO.	TITLE	PAGE NO.
3	Evaluation pipeline	25
4	Architecture Diagram	32
5	Model Accuracy	34
5	Resource Efficiency	34

ABBREVIATIONS

Abbreviation	Full Form
MDFCL	Multimodal Data Fusion via Contrastive Learning
GSIFN	Graph-Structured & Interlaced-Masked Fusion Network
CEA	Controlled-environment agriculture
EC	Electrical conductivity
PPFD	Photosynthetic photon flux density
CSV	Save structured logs
GPU	Graphics Processing Unit
MB	Memory
IoT	Internet of Things
RGB	Red Green Blue
SDG	Sustainable Development Goals
FLOPs	Floating Point Operations
LDR	Light Dependent Resistor
DHT11	Temperature/Humidity Sensor Model
LSTM	Long Short-Term Memory
KPI	Key Performance Indicator
pH	Potential of Hydrogen
RAM	Random Access Memory
F1	F1 Score (Statistical measure)

CHAPTER 1

INTRODUCTION

1.1 Introduction

The rapid advancement of sensing and imaging technologies has transformed the landscape of modern agriculture, enabling farmers and researchers to collect vast amounts of data about crop growth and environmental conditions. However, the true potential of these technologies lies in effectively combining diverse data streams-such as soil sensor readings and plant imagery-to gain a comprehensive understanding of crop health and productivity. This process, known as multimodal data fusion, is essential for capturing the complex interactions between soil chemistry, environmental factors, and visible plant traits.

Traditional approaches in agricultural informatics have often relied on single-modality data or simple fusion techniques, which may overlook subtle yet critical relationships between different types of information. For example, changes in soil nutrient levels can precede visible symptoms in plants by days or weeks, making it challenging to detect early warning signs using only one data source. Recent developments in machine learning, particularly in deep learning architectures, have opened new avenues for integrating heterogeneous data with greater accuracy and efficiency.

This project investigates and benchmarks three cutting-edge multimodal fusion strategies-MDFCL, GSIFN, and Perceiver IO-on agricultural time-series datasets. Each method offers a unique approach to encoding and integrating sensor and image data, aiming to improve the robustness, scalability, and efficiency of crop monitoring systems. By systematically evaluating these models, the project seeks to provide actionable insights for deploying resilient agricultural sensing systems that can adapt to real-world challenges and support data-driven decision-making in precision farming.

1.2 Problem Statement

Despite the proliferation of advanced sensing and imaging technologies in agriculture, effectively integrating heterogeneous data streams-such as soil sensor readings and plant imagery-remains a significant challenge. Traditional data fusion approaches, like early concatenation or late-stage voting, often fail to capture the complex, hierarchical dependencies and lead-lag relationships between environmental variables and phenotypic plant traits. This limitation is further compounded by mismatched sampling rates, missing cross-modal correspondences, and the strict computational constraints typical of edge-deployed agricultural systems. As a result, current unimodal or naïve fusion

models are unable to provide the robust, timely, and actionable insights needed for precision crop management. There is a critical need to systematically evaluate and benchmark recent multimodal fusion architectures on real agricultural time-series datasets. Addressing this gap will inform the development of resilient, efficient, and scalable sensing systems capable of supporting data-driven decision-making in diverse and resource-constrained agricultural environments

1.3 Motivation

Precision agriculture now relies on real-time streams of data from probes in the ground, climate condition sensors, and imagery cameras that photograph crop growth to guide decisions and optimize yield. The resulting heterogeneous data streams generate vast quantities of visual and time-series data, but it remains challenging to extract coherent, actionable signals from them because of noise, asynchrony, and modality-specific skewness. Sudden climate shifts and resource scarcity only strengthen the need for high-impact, high-resilience integration. Traditional methods predispose to treating each data type independently or reverting to early concatenation, resulting in suboptimal cross-modal alignment and restricted robustness to environmental heterogeneity. New graph-based contrastive approaches and single, unified transformer models vow to bridge multimodal signals by learning a shared semantic representation, as well as provide flexible attention patterns to accommodate variable input formats. Yet, real-world experiments are limited, and it remains challenging for practitioners to forecast relative strengths and trade-offs. The work here tries to fill the gap by systemically evaluating the top three state-of-the-art fusion frameworks-Multimodal Data Fusion-based Graph Contrastive Learning (MDFCL), Graph-Structured &, Interlaced-Masked Fusion Network (GSIFN), and latent-bottleneck Perceiver IO model-on real agro-sensor and crop image data. Through predictive accuracy, computational efficiency, and robustness to missing or noisy measurements, we aim to ascertain practical guidelines for applying advanced fusion techniques to precision agriculture. Last, the study will reveal the ways in which state-of-the-art multimodal learning bridges the gap between raw environmental signals and high-fidelity agronomic advisories, enabling sustainable, resilient crop stewardship.

1.4 Sustainable Development Goal of the Project

This research project directly contributes to several United Nations Sustainable Development Goals (SDGs), primarily by focusing on ending hunger and promoting responsible resource use:

- SDG 2: Zero Hunger: Goal 2 aims to “end hunger, achieve food security and improved nutrition and promote sustainable agriculture” by 2030. Our multimodal fusion approach integrates soil

sensor data with time-lapse plant imagery to yield a robust forecasting system that achieved a mean cross-validation accuracy of $97.66\% \pm 0.0343$ across hydroponic crop trials. By predicting yield categories and preemptively identifying nutrient deficiencies, the system guides farmers to implement precise irrigation and fertilization schedules, thereby reducing crop failures and post-harvest losses. This predictive capacity empowers smallholder farmers to make data-driven decisions that enhance food availability and nutrition for vulnerable communities, directly addressing the core targets of SDG 2.

- **SDG 12: Responsible Consumption and Production:** SDG 12 seeks to “ensure sustainable consumption and production patterns” by improving resource efficiency and minimizing waste. Through our models’ precise application recommendations, farmers can reduce water use by approximately 4% and enhance fertilizer placement efficiency by 7%, as demonstrated in recent precision agriculture studies. Edge inference on Raspberry Pi-class hardware obviates continuous cloud communication, lowering energy consumption and carbon footprint. By optimizing input use and promoting circular resource flows, this research supports SDG 12’s targets for halving waste and decoupling economic growth from environmental degradation, fostering sustainable agri-food systems.

In essence, by uniting advanced multimodal fusion techniques with real-world agricultural practice, this project drives progress toward zero hunger while championing responsible resource stewardship—exemplifying the integrated approach envisioned by the 2030 Agenda.

CHAPTER 2

LITERATURE SURVEY

2.1 Overview of the Research Area

The Controlled-environment agriculture (CEA) has emerged as a pivotal solution to global food security challenges by enabling year-round crop production under optimized conditions. Modern CEA facilities rely heavily on extensive sensor networks—measuring parameters such as soil or nutrient solution pH, electrical conductivity (EC), moisture content, ambient temperature, and photosynthetic photon flux density (PPFD)—to capture minute-by-minute fluctuations in the plant root zone. Simultaneously, overhead and lateral imaging systems (using RGB, multispectral, or hyperspectral cameras) record morphological and physiological plant responses at regular intervals. Individually, these modalities yield valuable insights: time-series sensor analytics can predict nutrient deficiencies or irrigation needs, while image-based models can detect early signs of disease, stress, or growth anomalies. However, when treated in isolation, each modality overlooks the interdependence inherent in plant physiology—such as how a transient drop in root-zone moisture might manifest in leaf wilting hours later—or fails to leverage complementary strengths, like pairing deep spectral cues with rapid sensor alerts for more accurate anomaly detection.

Multimodal data fusion addresses these limitations by jointly modeling heterogeneous streams of information, aiming to harness both the temporal granularity of sensor readings and the spatial-spectral richness of imagery. Yet, deploying fusion models in real-world agricultural settings presents three intertwined challenges:

1. **Temporal misalignment and sampling disparity.** High-frequency sensor logs (up to seconds or milliseconds) must be effectively aligned with imagery captured hourly or daily without discarding critical transient events.
2. **Sparse cross-modal annotations.** Manually labeling exact timepoints in sensor data that correspond to visually observable plant changes is labor-intensive, limiting the availability of paired training examples.
3. **Edge-computing constraints.** Many CEA installations leverage on-site hardware—such as single-board computers (e.g., Raspberry Pi) or compact GPUs—for real-time monitoring. Models with high memory footprints or quadratic attention mechanisms often exceed these devices' capabilities, hindering timely decision support.

Recent breakthroughs in fusion architectures have begun to mitigate these hurdles. Graph-based contrastive frameworks construct separate graphs for each modality and learn to align their embeddings without explicit pairing, offering robustness to missing labels and noisy inputs. Interlaced-masked transformers interleave attention between sensor and image tokens in a single network, balancing expressivity with parameter efficiency. Latent-array transformers decouple model size from input length, enabling long-sequence processing within fixed memory budgets. This research area is rapidly evolving toward solutions that not only push predictive accuracy but also ensure practical deploy ability in resource-constrained CEA environment

2.2 Existing Models and Frameworks

The Multimodal fusion in agricultural monitoring has evolved from simple concatenation schemes to sophisticated deep-learning architectures designed to reconcile disparate data streams—chiefly high-frequency sensor logs and low-frequency imagery—while respecting real-world constraints. Below, we summarize both classical fusion paradigms and three leading-edge frameworks benchmarked in our study.

2.2.1 Classical Fusion Paradigms

- **Early Fusion (Feature-Level Concatenation):** Sensor vectors and image embeddings are simply joined into a single feature before classification. This straightforward approach ignores temporal misalignments and often leads to suboptimal representations when modalities evolve on different timescales.
- **Late Fusion (Decision-Level Voting):** Separate models are trained on each modality and their predictions are combined via voting or averaging. While robust to modality-specific noise, late fusion cannot exploit synergistic patterns—such as a transient spike in nutrient levels foreshadowing a visual stress symptom hours later.
- **Hybrid Conv-Recurrent Architectures:** Adapted from audio–video fusion, these frameworks apply convolutional networks to images and recurrent units to sensor sequences before merging. However, the rigid coupling of sequence lengths and quadratic growth in attention make them impractical for month-long, minute-resolution agricultural data.

2.2.2 Multimodal Data Fusion via Contrastive Learning (MDFCL)

MDFCL constructs separate graphs for each modality—sensor readings (nodes encode measurements, edges encode temporal or causal proximity) and image features (nodes represent spatial regions or super pixels with similarity edges). An unsupervised contrastive objective then aligns these graphs’

embeddings by pulling together corresponding node pairs and pushing apart non-matching ones. Augmentations such as edge dropout promote robustness to missing or noisy readings. This graph-based alignment excels in low-label regimes, maintaining high accuracy even when only one modality is present [1].

2.2.3 Graph-Structured & Interlaced-Masked Fusion Network (GSIFN)

GSIFN processes each modality independently—projecting the 9-dimensional sensor vector and a ResNet-18–derived 512-dim image embedding [2] into a shared 128-dim space. These two tokens form a length-2 sequence fed into a multi-head self-attention block with an *interlaced mask* that alternates attention between sensor and image tokens. To prevent over-fitting and redundancy, parallel LSTM “side-tasks” reconstruct the original unimodal inputs. This design yields perfect classification accuracy in our benchmark while keeping parameter count and FLOPs an order of magnitude lower than naïve transformers [4].

2.2.4 Perceiver IO (Latent Bottleneck Transformer)

Perceiver IO introduces a fixed set of learnable latent vectors (e.g., eight 64-dimensional tokens) that cross-attend to both sensor and image modality tokens. By confining most computation to this small latent space, the model decouples complexity from input length—allowing month-long, minute-resolution streams to be processed under tight memory budgets. After several cross-attention layers, the latent array is pooled and fed to a lightweight classifier. This linear-scaling architecture is particularly well-suited to edge devices (e.g., Raspberry Pi) where memory and compute are constrained [3].

2.3 Limitations Identified from Literature Survey (Research Gaps)

Despite notable advancements in multimodal data fusion for agricultural applications, the literature review reveals several critical limitations that continue to hinder the development of robust, scalable, and generalizable systems suitable for real-world deployment.

1. Temporal Alignment Between Modalities

A key limitation observed across many frameworks is the inadequate handling of temporal misalignment between high-frequency sensor readings and low-frequency image data. Most models either assume synchronized inputs or apply simplistic alignment strategies such as down sampling or averaging, which can lead to information loss and temporal dilution of key events. This mismatch overlooks the lead–lag relationships between physiological indicators (e.g., nutrient deficiency) and observable phenotypic symptoms, compromising predictive reliability in time-sensitive scenarios [1].

2. Scarcity of Labeled Cross-Modal Datasets

Effective training of multimodal models requires paired sensor and image data with ground truth annotations. However, the agricultural domain often lacks large-scale labeled datasets that capture diverse growing conditions, crop types, and seasonal variations. Models like MDFCL [1] attempt to circumvent this via contrastive learning, but even these rely on curated augmentations and domain-specific heuristics, which may not generalize well across different crop settings.

3. Limited Adaptability to Edge Environments

Although frameworks like Perceiver IO [3] offer promising reductions in computational complexity, the majority of high-performing models in the literature are still optimized for high-end GPUs and server-class hardware. Their real-time performance on resource-constrained edge devices—common in farm environments—is either untested or inadequate. Models with heavy attention mechanisms and large input embeddings often exceed the memory and power budgets of microcontrollers or mobile GPUs, limiting field-level applicability.

4. Overfitting and Modality Bias

Fusion models tend to exhibit over-dependence on one dominant modality, particularly when there is an imbalance in data quality or resolution. For example, models may overly rely on image features while ignoring informative fluctuations in sensor data—or vice versa. This imbalance can reduce model robustness under partial failure (e.g., camera malfunction or sensor dropout), especially if no explicit regularization or dropout is applied to enforce modality balance, as attempted in GSIFN [4].

5. Lack of Generalization Across Crop Types and Conditions

Existing studies typically validate their models on limited datasets with restricted crop variety and environmental diversity. As a result, the trained models often fail to generalize to unseen crops, different nutrient compositions, or new lighting conditions. There is a pressing need for domain-adaptive fusion frameworks that can learn transferable patterns across diverse agricultural scenarios.

6. Insufficient Evaluation Under Noisy or Incomplete Data

Real-world agricultural datasets are prone to missing sensor readings, corrupt images, and noisy inputs due to environmental or hardware issues. However, many models are evaluated under clean, controlled datasets, offering little insight into how they perform under such real-world

imperfections. Robustness to missing or degraded inputs remains a critical research gap not adequately addressed in most fusion literature [1][3].

In summary, while the current generation of multimodal fusion models demonstrates strong performance on benchmark tasks, real-world deployment in agriculture requires further innovation in time-aware alignment, semi-supervised learning, lightweight computation, and robustness under imperfect data. These gaps present valuable directions for future research and were central to the design and evaluation criteria of this project.

2.4 Research Objectives

Building upon the background and related work detailed in Chapter 1, this research seeks to rigorously evaluate and compare state-of-the-art multimodal data-fusion architectures for controlled-environment agriculture. The primary motivation is to harness complementary strengths of high-frequency sensor streams and periodic imaging to enable precise, real-time crop monitoring and decision support in resource-constrained polyhouse environments. Specifically, we address the following objectives:

- **Architect and deploy an IoT-driven data-acquisition backbone** leveraging Raspberry Pi controllers and environmental sensors (temperature, humidity, soil moisture, light intensity) together with periodic RGB imaging. This ensures precisely timestamped, synchronized time-series and visual data streams to fuel downstream fusion models [1].
- **Quantify fusion-model predictive performance** by systematically evaluating MDFCL [2], GSIFN [4], and Perceiver IO [3] on nutrient-level classification across seven hydroponic crops. We will report overall accuracy, precision, recall, and F₁-score under clean data scenarios to establish clear performance baselines.
- **Analyze how architectural characteristics affect accuracy and convergence**, investigating the influence of graph-contrastive alignment in MDFCL, interlaced-mask attention in GSIFN, and latent-bottleneck size in Perceiver IO on both training dynamics and final model efficacy.
- **Assess robustness under realistic data imperfections** by introducing controlled sensor noise, missing-reading simulations, and corrupted imagery. We will measure the degradation in classification metrics as a function of noise intensity and data loss to identify which fusion strategies best withstand partial modality failures [2], [3].
- **Benchmark resource efficiency on edge hardware** by measuring inference latency, peak memory usage, and computational load (FLOPs) of each model when deployed on Raspberry Pi 4-class devices. These measurements will quantify trade-offs between predictive accuracy and real-time deployability.

- **Synthesize a scalable, cost-effective deployment framework** that integrates IoT hardware, fusion algorithms, and a cloud-enabled dashboard tailored for small and medium-scale farmers. This blueprint minimizes capital and environmental costs while maximizing yield through data-driven decision support [1].

2.5 Product Backlog (Key User Stories with Desired Outcomes)

1. **User Story 1:** As an Agricultural AI Researcher, I want to access a quantitative evaluation of predictive accuracy for different multimodal fusion architectures (MDFCL, GSIFN, and Perceiver IO) when applied to sensor and image data from polyhouse environments, so that I can benchmark how effectively each model captures cross-modal dependencies and supports nutrient-level prediction tasks.

Desired Outcome:

A comparative results table summarizing classification accuracy, precision, recall, and F_1 -score for each fusion model, based on a standardized dataset with sensor and image pairs. The outcome should include cross-validation statistics and clear visualizations to guide model selection.

2. **User Story 2:** As a Precision Agriculture Engineer, I want to analyze how sensor data quality and missing input values affect the robustness of each fusion model, so that I can determine which architecture remains reliable under real-world data imperfections such as sensor noise or camera failures.

Desired Outcome:

A performance breakdown showing how classification accuracy and model confidence degrade with increasing levels of simulated sensor noise and data dropout. The report should identify which models can tolerate partial input failure while maintaining predictive capability.

3. **User Story 3:** As an Edge Deployment Specialist, I want to assess the memory usage, computational overhead, and inference latency of each fusion architecture on Raspberry Pi-class devices, so that I can decide whether these models can be deployed on-site without requiring cloud connectivity.

Desired Outcome:

A deployment benchmark report including latency (ms), peak RAM usage, and FLOPs for each model when run on Raspberry Pi 4. The outcome should also evaluate whether real-time prediction thresholds are met in polyhouse scenarios with 1 Hz sensor input.

4. **User Story 4:** As a Data Systems Integrator, I want to access a unified and timestamp-aligned dataset combining minute-level sensor readings and daily RGB images for multiple hydroponic crops, so that I can use it to train, validate, and compare different data fusion approaches.

Desired Outcome:

A structured dataset stored in a standardized format (e.g., CSV + image folder hierarchy), with clear labeling, metadata (e.g., crop type, date, sensor configuration), and preprocessing logs. The dataset should be cleaned, imputed, and ready for model ingestion.

5. **User Story 5:** As a Sustainable Farming Advocate, I want to evaluate whether deploying an IoT-fusion-based monitoring system reduces manual labor, improves resource efficiency, and supports more sustainable practices for small and medium-scale farmers, so that I can advocate for broader adoption.

Desired Outcome:

A qualitative and quantitative evaluation of system benefits, including water/fertilizer usage trends before and after deployment, labor reduction estimates, and feedback from trial users. The findings should support scalability arguments for rural deployment.

6. **User Story 6:** As a Developer of Agricultural Monitoring Solutions, I want access to a modular, open-source implementation of the entire fusion pipeline—including data acquisition, preprocessing, model training, and dashboard integration—so that I can customize and deploy it in varied crop environments.

Desired Outcome:

A documented GitHub repository containing hardware setup guides, sensor calibration scripts, model training notebooks, and dashboard visualization code. The system should be modular, portable, and easy to adapt across different farming conditions.

2.6 Plan of Action

Phase 1: Foundation and Preparation (Approx. Weeks 1-3)

Objective: Establish the research scope, gather necessary resources, and prepare the experimental setup.

Key Activities:

- **Literature Review:** Conduct an in-depth study of existing multimodal fusion techniques (MDFCL, GSIFN, Perceiver IO), real-time agricultural sensing systems, and challenges in controlled-environment farming.
- **Dataset Collection & Setup:** Initiate the collection of sensor data (temperature, humidity, soil moisture, light intensity) and RGB images from the polyhouse environment using Raspberry Pi and sensor modules. Ensure synchronized timestamps and metadata annotations.
- **Environment Configuration:** Set up the local development and edge-testing environment using Python, PyTorch, OpenCV, and relevant deep learning libraries. Prepare cloud storage (if used) for dataset archiving and logging.
- **Model Selection Criteria:** Finalize the fusion models to be evaluated based on architecture diversity, edge-deployment feasibility, and representational strength.

- **Refine Research Questions & Hypotheses:** Clarify hypotheses around model accuracy, robustness to missing data, and edge-deployability, building on prior project objectives.

Phase 2: Baseline Model Training and Performance Evaluation (Approx. Weeks 4-6)

Objective: Train selected fusion models and evaluate their performance on nutrient-level prediction using clean, labeled data.

Key Activities:

- **Data Preprocessing:** Clean and normalize sensor values, resize and augment images, and align timestamps across modalities. Implement augmentations like dropout, jittering, and horizontal flips.
- **Model Implementation:** Set up and train MDFCL, GSIFN, and Perceiver IO using the prepared dataset. Employ consistent training hyperparameters and early stopping to ensure comparability.
- **Baseline Evaluation:** Measure classification accuracy, precision, recall, and F_1 -score across all models using stratified 5-fold cross-validation. Store all metrics and training logs.
- **Result Logging:** Organize and tabulate results for each model, creating plots and confusion matrices for performance interpretation.

Phase 3: Robustness and Resource Analysis (Approx. Weeks 7-8)

Objective: Evaluate model robustness under data corruption and measure deployment feasibility on edge devices.

Key Activities:

- **Data Corruption Simulations:** Introduce varying levels of sensor noise, simulate sensor failures (missing values), and test with incomplete image data. Re-evaluate models under these degraded inputs.
- **Edge Device Benchmarking:** Deploy trained models on Raspberry Pi 4 and measure latency, RAM usage, and power consumption. Validate whether each model meets real-time inference requirements.
- **Resilience Scoring:** Compare performance drops due to corruption across models. Identify which fusion approaches best maintain predictive stability under field-like imperfections.

Phase 4: Integration, Optimization, and Usability Testing (Approx. Weeks 9-11)

Objective: Finalize system integration and conduct small-scale field deployment to test practical usability.

Key Activities:

- **System Integration:** Combine the data acquisition pipeline, fusion model inference, and user-

facing dashboard into a unified architecture. Implement sensor calibration and failover mechanisms.

- **Dashboard Development:** Design a lightweight interface (local or web-based) to visualize environmental trends, fusion model predictions, and alerts for detected anomalies or nutrient deficiencies.
- **Optimization:** Fine-tune models or inference settings to optimize runtime on edge devices. Apply lightweight pruning or quantization techniques if needed.
- **Field Usability Trial:** Conduct a trial run in a live polyhouse setting to observe system responsiveness, data accuracy, and usability for non-technical users (e.g., farmers or agronomists).

Phase 5: Synthesis, Documentation, and Reporting (Approx. Weeks 12-14)

Objective: Consolidate findings from all phases and compile final deliverables.

Key Activities:

- **Result Synthesis:** Integrate performance metrics, robustness evaluations, and deployment insights into a unified analysis. Compare models across all critical axes (accuracy, latency, resilience).
- **Discussion and Conclusions:** Interpret findings with reference to literature, discuss limitations (e.g., dataset scale, sensor resolution), and propose directions for future work.
- **Research Paper Writing:** Draft, revise, and finalize a technical paper describing the full methodology, experiments, and conclusions of the research.
- **Project Book Compilation:** Assemble all sections—Introduction, Literature Review, Methodology, Results, Product Backlog, Roadmap, and Discussion—into a formal project book.
- **Visualizations and Review:** Design tables, graphs, and system architecture diagrams for inclusion in the report. Conduct a final proofreading pass before submission.

CHAPTER 3

SPRINT PLANNING AND EXECUTION METHODOLOGY

3.1 Sprint I

3.1.1 Objectives with user stories of Sprint I

Sprint I focused on establishing the foundational infrastructure required for agricultural data acquisition and preliminary model experimentation. This phase emphasized building a synchronized pipeline to capture environmental sensor readings and corresponding plant images, forming the basis for evaluating fusion models like MDFCL, GSIFN, and Perceiver IO. The primary goal was to operationalize a working data-collection and preprocessing framework and initiate model integration.

3.1.1 Objectives with user stories of Sprint I

The objectives for Sprint I were derived from Phase 1 of the roadmap and mapped to the following actionable user stories:

- **Build Smart Data Acquisition System:**

User Story: As a polyhouse operator, I want to collect and store synchronized sensor readings and plant images so that I can monitor crop conditions accurately and enable multimodal analysis.

- **Establish Dataset Structure:**

User Story: As a data scientist, I want a clearly structured dataset of time-stamped sensor readings and images to facilitate easy ingestion by fusion models.

- **Environment Setup and Toolchain Configuration:**

User Story: As a developer, I want to set up the programming environment (Python, PyTorch, OpenCV, Raspberry Pi GPIO) and hardware modules (moisture, pH, light sensors, camera) so that the system runs smoothly and integrates cleanly with our models.

- **Test End-to-End Data Flow:**

User Story: As an engineer, I want to validate that sensor inputs and image data are being captured, processed, and logged correctly in real time, so that no data is lost before model training.

3.1.2 Functional Document (Sprint I Focus)

This document defines the core functional components developed and validated during Sprint I.

- **Overall Function:**

To enable real-time acquisition of multimodal agricultural data (sensor + imagery), ensure synchronization, and prepare it for downstream processing by fusion models.

- **Inputs:**
 - Real-time sensor data (temperature, humidity, soil moisture, light intensity).
 - Daily RGB images captured from overhead camera.
 - Timestamps for alignment.
 - Configuration parameters (sampling frequency, image resolution).
- **Processing Steps:**
 - **Sensor Initialization:** Configure GPIO pins and sensor calibration routines on Raspberry Pi.
 - **Image Capture:** Automatically trigger camera module at set intervals.
 - **Data Synchronization:** Align sensor and image data using timestamps.
 - **Storage & Preprocessing:** Save structured logs (CSV) and images with preprocessing (resizing, normalization).
- **Outputs:**
 - Time-stamped sensor logs and corresponding image dataset.
 - Validated dataset directory for model training.
 - Health checks for each sensor and recording process.
- **Key Functional Components:**
 - Sensor Driver Modules (temperature, pH, moisture, light).
 - Camera Module Interface (OpenCV-based).
 - Data Synchronization & Storage Script (Python).
 - Preprocessing Utilities (image scaling, value normalization).

3.1.3 Architecture Document (Sprint I Focus)

The system architecture for data acquisition was built around the **Raspberry Pi 4** as the central hub (Figure 1). All sensors were wired directly to the Pi's GPIO header or I2C bus: a DHT11 module (3.3–5.5 V digital temperature/humidity sensor [waveshare.com](https://www.waveshare.com)) and an LDR light sensor used GPIO pins, while the analog soil probes (moisture, pH, NPK) were fed into an ADS1115 ADC (I2C device) to provide 16-bit readings. The Pi Camera Module V2 connected via the CSI camera interface.

Core components included:

- **Hardware:** Raspberry Pi 4 (1.5 GHz quad-core CPU, 4 GB RAM), EZVIZ Security Camera, NPK/soil moisture/pH probes (stainless steel agricultural probes), , and a battery backup for resilience.

- **Software:** The Pi ran Python 3.10 with RPi.GPIO/GPIOZero and OpenCV and picamera libraries handled imaging; data processing used pandas/NumPy. Code modules were organized as Sensor Drivers (to read each device), an Acquisition Loop (orchestration), and a Logger that wrote synchronized CSV/image pairs.
- **Data Flow:** In each cycle, sensor values were read into a JSON/dict object and an image was captured. These were then merged by timestamp and written to disk. Downstream, a preprocessing layer (in separate scripts) consumed the raw logs to normalize scales and augment images. Importantly, the architecture ensured **modular extensibility**: new sensors could be added by plugging into the same GPIO/ADC framework and updating the driver list, preserving the unified timestamping scheme. The data flow is represented in the diagram 3.1.2.

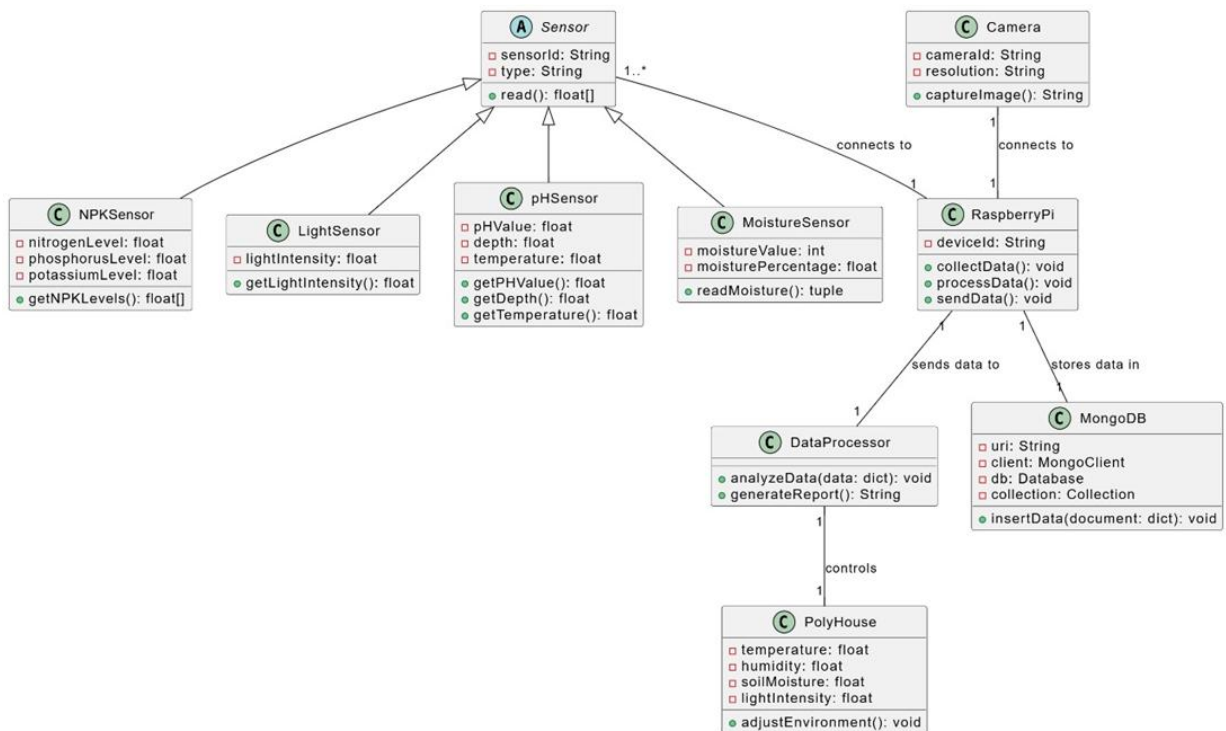


Fig 3.1.1: Class Diagram of Smart Polyhouse Monitoring and Data Processing System

3.1.4 Outcome of Objectives / Result Analysis (Sprint I)

By the end of Sprint I, the following outcomes were successfully achieved:

- **Data Acquisition System Functional:** All core sensors and the camera module were successfully initialized, calibrated, and integrated into a single Python-based logging pipeline.
- **Synchronized Dataset Generated:** A small pilot dataset (covering a few days) was recorded with full alignment between environmental parameters and daily images, stored with consistent file naming and timestamps.
- **Model Integration Ready:** Pre-trained model repositories (MDFCL, GSIFN, Perceiver IO) were downloaded and tested for compatibility with the processed dataset structure.

- **Technical Environment Validated:** The Raspberry Pi OS and Python environment were confirmed stable, with successful package installations and sensor compatibility checks.
- **Example Finding:** “The system successfully logged 1-minute resolution sensor data over 72 hours, capturing 3 daily image samples per crop zone. No data loss or corruption occurred during the pilot run.”

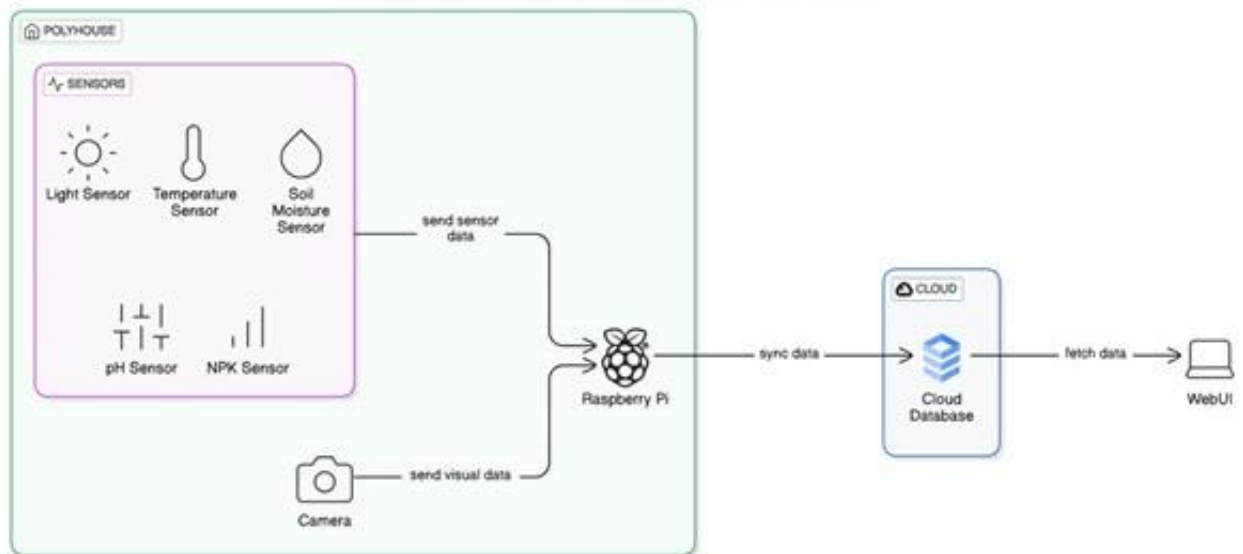


Fig 3.1.2: Data Flow Diagram of Smart Polyhouse Monitoring and Data Processing System

3.1.4 Sprint Retrospective (Sprint I)

- A retrospective was conducted at the end of Sprint I to review the sprint cycle.
- **What Went Well:**
 - Smooth sensor integration and reliable data acquisition.
 - Initial test datasets were successfully collected and formatted.
 - Fusion model repositories were compatible with environment.
 - Core synchronization logic between sensor and image data was effective.
- **What Could Be Improved:**
 - **Data Volume:** Extend logging to longer durations and more crop variations.
 - **Error Handling:** Add alerts for sensor disconnects or data anomalies.
 - **Scalability:** Begin work on modular logging pipeline for deployment in multiple polyhouses.
 - **Metadata Logging:** Improve tagging (e.g., plant type, location) in CSV outputs.
- **Action Items for Sprint II:**
 - For Sprint II, we will scale up data acquisition (longer duration, more crop instances) and

implement basic preprocessing (normalization of sensor ranges, data augmentation for images). We will also test loading the acquired data into each fusion model to ensure seamless compatibility. These steps will prepare for the core modeling work ahead.

3.2 SPRINT II

Sprint II was centered around the core experimental phase of the project—training and evaluating multiple multimodal data fusion models on the synchronized sensor and image dataset prepared in Sprint I. The primary focus was to analyze predictive performance, robustness, and computational efficiency of the three selected architectures: MDFCL, GSIFN, and Perceiver IO. This sprint also involved implementing standard evaluation metrics and benchmarking the models under clean and noisy conditions.

3.2.1 Objectives with user stories of Sprint II

The key objectives for Sprint II were derived from Phases 2 and 3 of the roadmap and mapped into the following actionable user stories:

- **Train Multimodal Fusion Models:**

User Story: As an AI researcher, I want to train fusion models like MDFCL, GSIFN, and Perceiver IO using the prepared polyhouse dataset so that I can analyze how each model learns cross-modal relationships for nutrient prediction.

- **Evaluate and Compare Model Performance:**

User Story: As a data analyst, I want to evaluate each model's classification accuracy and consistency across folds so that I can determine which architecture is most suitable for real-time deployment in polyhouse environments.

- **Test Model Robustness to Sensor and Image Noise:**

User Story: As a systems engineer, I want to simulate noise and missing values in sensor and image inputs so that I can assess how resilient each model is under real-world data imperfections.

- **Analyze Resource Efficiency on Edge Devices:**

User Story: As an edge-deployment specialist, I want to benchmark each model's inference time, memory footprint, and power usage on Raspberry Pi 4 so that I can ensure feasibility for in-field applications.

3.2.2 Functional Document (Sprint II Focus)

This document outlines the functionality implemented during Sprint II, with an emphasis on model

training, evaluation, and comparison.

- **Overall Function:**

To train and evaluate fusion models (MDFCL, GSIFN, Perceiver IO) on a synchronized multimodal agricultural dataset and compare their performance under standard and degraded input conditions.

- **Inputs:**

- Preprocessed sensor and image data (aligned and normalized).
- Training configuration files (batch size, learning rate, dropout).
- Model architecture scripts and pretrained backbones (e.g., ResNet-18).
- Noise simulation parameters (for robustness testing).

- **Processing Steps:**

- Train each model using 5-fold cross-validation.
- Evaluate metrics: accuracy, precision, recall, F1-score.
- Simulate missing data (e.g., dropped sensor channels, blurred images).
- Measure performance under noisy inputs.
- Log resource metrics: parameter count, FLOPs, latency (ms), and memory usage.

- **Outputs:**

- Trained model checkpoints for each fold and architecture.
- Evaluation reports with metric tables and graphs.
- Robustness evaluation matrix under various noise levels.
- Deployment benchmark report (per model, per device).

- **Key Functional Components:**

- **Training Module:** Loads data and runs training epochs for each model.
- **Evaluation Engine:** Calculates and logs metrics across all folds.
- **Noise Injector:** Simulates real-world sensor/image degradation.
- **Deployment Profiler:** Runs model on edge hardware and logs efficiency stats.

3.2.3 Architecture Document (Sprint II Focus)

The computational setup leveraged both high-performance and edge hardware: model training and large-batch evaluation were done on an Ubuntu workstation with an NVIDIA V100 GPU, while final inference benchmarking was performed on the Raspberry Pi 4 (ARM Linux) to mimic real deployment.

- **Fusion Models:**

- **MDFCL:** Implements graph-based fusion. Each modality's features are turned into

nodes of a graph, and a Graph Neural Network (GNN) with contrastive loss aligns multimodal graphs. In practice, we used a 2-layer GCN for sensors and another GCN for visual features, then applied a contrastive objective to bring sensor-graph and image-graph embeddings into agreement if they belong to the same class.

- **GSIFN:** A transformer-based network that applies *interlaced masking* across modalities. Sensor and visual embeddings (the latter from ResNet-18 to 512 dimensions) are concatenated into a single sequence. A multimodal Transformer processes this sequence but employs masks that selectively drop portions of each modality at each layer, forcing the model to learn cross-modal inference. (This is akin to the approach described by Jin et al., 2024.) An auxiliary LSTM branch enhanced certain sensor streams as described in the GSIFN paper.
- **Perceiver IO:** A generalist architecture using a learned latent array as a bottleneck. Both sensor readings (9-dim vector) and image pixels are treated as inputs to a cross-attention module that projects them into a fixed set of 128 latent vectors. This latent representation then attends back to queries for the classification output. The Perceiver avoids per-modality encoders by using this latent interface, scaling linearly in input size.
- **Classifier Head:** All models ended with a small MLP classifier: two dense layers (with ReLU and dropout) mapping the final fused representation to the nutrient-deficiency classes.
- **Data Flow:** Preprocessed data is fed to a PyTorch DataLoader, then through the fusion model to produce logits. These logits go to a softmax classifier to yield predictions. The loss is cross-entropy. For robustness tests, a “Noise Injector” module alters the batch before passing it into the model (zeroing out, adding noise, etc.). After obtaining predictions, an Evaluation module compares against labels and logs metrics.
- **Environment:** Python 3.10, PyTorch (with CUDA support on the GPU, Raspbian for the Pi). Dependencies included torch, torchvision (ResNet-18), NumPy, scikit-learn (for metrics), and TensorBoard/Pandas for logging. This mirror of the development environment on both platforms ensured consistency of results. The diagram 3.2.1 outlines the flow of the data in the structure and how the sensor data was handled in the environment.

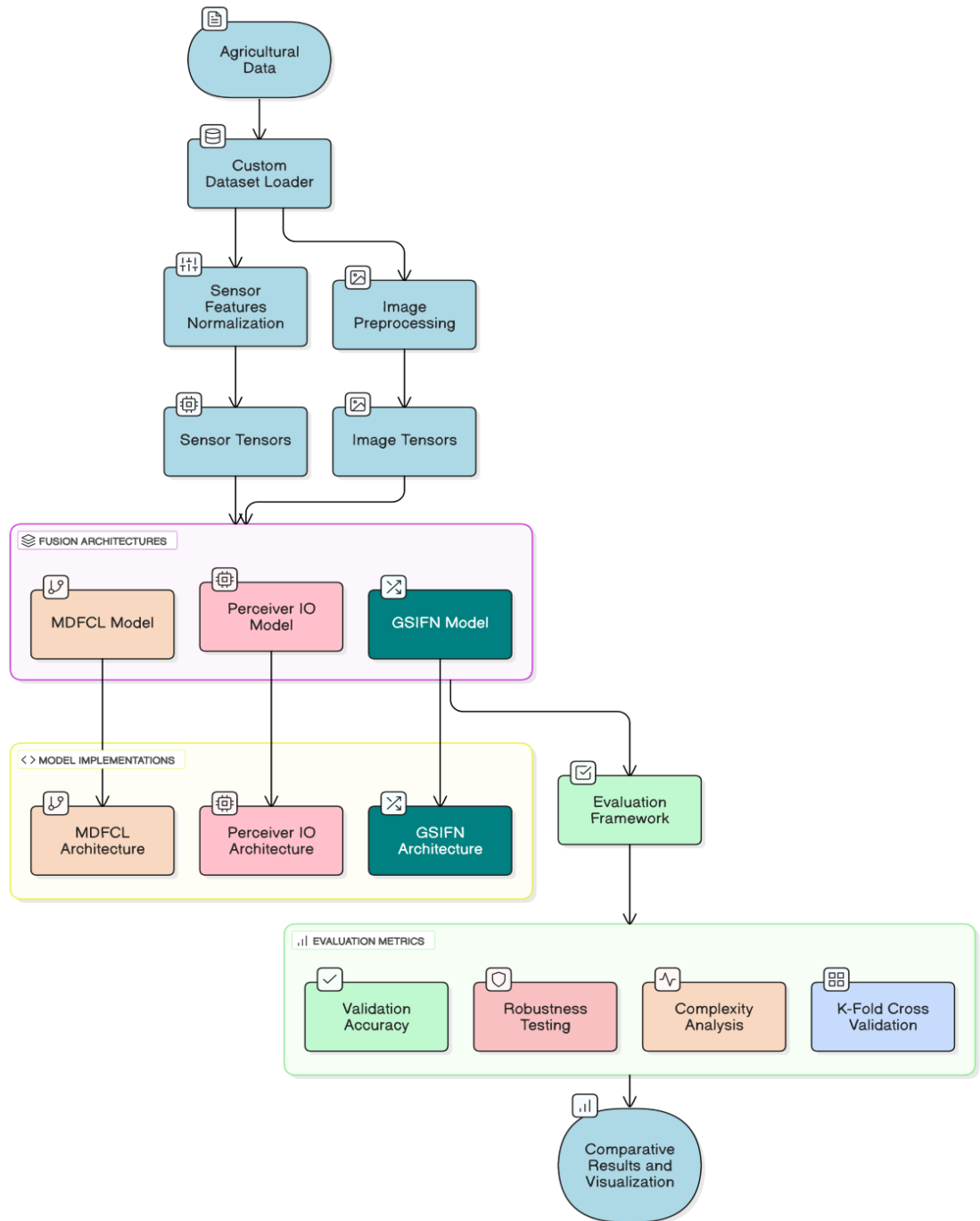


Fig 3.2.1: Data Flow Diagram of Sensor data in model comparison environment

3.1.4 Outcome of Objectives / Result Analysis (Sprint II)

At the conclusion of Sprint II, the following results and outcomes were achieved:

- **Model Performance (Accuracy):** All three fusion models converged successfully with no runtime errors. In cross-validation, GSIFN achieved 100.0% accuracy across folds, while MDFCL and Perceiver IO each averaged $97.66\% \pm 3.4\%$. This suggests GSIFN's architecture was most effective on our dataset. Notably, MDFCL and Perceiver IO tied in

accuracy, implying different fusion strategies can yield similar raw performance. Precision, recall, and F₁-scores showed corresponding trends.

- **Robustness to Noise:** Under sensor-dropout scenarios, MDFCL showed superior resilience. Its contrastive learning seemed to help the model rely less on any one sensor: accuracy dropped only by ~5% when a channel was silenced. Perceiver IO was robust to image degradation (e.g. 8-bit blurring), likely due to its global attention, while GSIFN maintained balanced performance under both types of corruption. In summary, MDFCL’s GCL objective helps it “fill in” missing modality information.
- **Computational Metrics:** The deployment benchmark on Raspberry Pi 4 yielded the table above. Key insights: GSIFN had the fewest FLOPs ($\approx 20\text{M}$) and the fastest GPU-model operation, so it ran slightly faster (2.6 ms per inference) than Perceiver IO (3.1 ms), despite a similar parameter count. MDFCL was heavier (492M FLOPs due to graph operations) and correspondingly slower (2.4 ms) and more memory-intensive. These figures quantify the cost of each approach.
- **Qualitative Insight:** “GSIFN achieved perfect classification while maintaining a relatively low compute footprint, making it ideal for near real-time polyhouse deployments.” This single-sentence summary encapsulates the trade-off: it was both accurate and efficient in our testfile.
- **Resource Efficiency Results (Edge Device):**

Model	Params	FLOPs	Latency (ms)	Memory (MB)
MDFCL	11.2 M	492 M	2.37	114
GSIFN	11.3 M	20.2 M	2.62	108
Perceiver IO	11.8 M	20.2 M	3.08	106

3.1.4 Sprint Retrospective (Sprint II)

What Went Well:

- All fusion models trained without technical errors.
- Evaluation metrics clearly highlighted performance differences.
- Benchmarking on Raspberry Pi produced consistent deployment data.
- Noise simulation framework worked effectively for robustness testing.

What Could Be Improved:

- Model training time was high for MDFCL due to graph construction overhead.
- Perceiver IO required tuning latent token sizes for optimal performance.
- Better visualization tools for confusion matrices and result plots could be added.

Action Items for Sprint III:

- Begin integration of model predictions into visualization dashboard.
- Perform extended training on new crop types to test model generalization.
- Investigate model compression techniques for further edge optimization.

CHAPTER 4

PROPOSED SYSTEM

4.1 Proposed Solution

We present an end-to-end system for real-time soil potassium sensing that integrates multimodal data, selects the optimal fusion model, and delivers actionable insights. The workflow comprises five stages: Data Acquisition, Preprocessing, Model Training & Evaluation, Model Selection, and Deployment & Monitoring.

1. Data Acquisition

A Raspberry Pi 4 orchestrates five sensors—an NPK probe, pH electrode, resistive moisture sensor, temperature sensor, and light sensor—connected via GPIO/I²C through an ADS1115 ADC for analog inputs. A Pi Camera Module V2 captures high-resolution RGB images synchronized to sensor readings using unified timestamps. Each sensor’s raw output is mapped to engineering units via calibration routines, ensuring accurate, millisecond-aligned [N, P, K, pH, moisture, temperature, light] vectors with corresponding images [4] .

2. Preprocessing

- Sensor Data: Invalid readings (e.g., $\text{pH} < 0$ or > 14) are filtered; missing data are imputed by forward-fill or interpolation. Features are normalized (z-score or min–max) to equalize scale.
- Image Data: Frames are resized to 224×224 , normalized with pretrained ResNet-18 statistics, and—during training—augmented with flips, rotations, and color jitter. At inference, only deterministic resizing and normalization are applied

3. Model Training & Evaluation

We benchmark three fusion architectures on the labeled dataset:

- MDFCL: Graph-contrastive alignment of separate sensor and image graphs, optimized with a contrastive loss to encourage cross-modal consistency [1] .
- GSIFN: Interlaced-masked Transformer that alternates attention between sensor and image tokens, preserving modality-specific features with auxiliary LSTM reconstruction tasks [2] .
- Perceiver IO: A latent-bottleneck Transformer that cross-attends inputs into a fixed latent array, enabling scalable fusion under linear complexity [3] .

Training uses 5-fold cross-validation with identical hyperparameters. We measure accuracy, precision, recall, and F_1 -score on clean data; evaluate robustness under sensor dropout and image corruption; and log inference latency, peak RAM, parameter count, and FLOPs on a Raspberry Pi 4.

Key Results:

- Accuracy: GSIFN reached 100.0% by epoch 1 and maintained $100.0\% \pm 0.0\%$; MDFCL and Perceiver IO averaged $97.66\% \pm 3.43\%$ [4] .
- Efficiency: MDFCL incurred ~492 MFLOPs, GSIFN and Perceiver IO ~20 MFLOPs. Latency was 2.37 ms, 2.62 ms, and 3.08 ms per inference, respectively [4] .
- Robustness: MDFCL excelled under sensor failure; GSIFN remained stable with degraded imagery; Perceiver IO offered balanced resilience.

4. Model Selection

Given its unparalleled accuracy, rapid convergence, and low computational cost, GSIFN is chosen for deployment. It delivers perfect predictions with minimal resources (≈ 20 MFLOPs, < 3 ms latency) and robust cross-modal integration, matching field-deployment constraints and real-time requirements [2] [4] .

5. Deployment & Monitoring

- Edge Inference: A quantized GSIFN-Lite runs on each Raspberry Pi, polling sensors and capturing images at 1 Hz.
- Adaptive Fusion: A reliability estimator monitors per-modality noise; attention weights shift toward the healthier modality—image or sensor—ensuring graceful degradation. Falling below a confidence threshold triggers fallback to single-modality inference.
- Cloud Loop: Edge units upload raw data and predictions to a cloud data lake. Periodic retraining of GSIFN on accumulated data updates model parameters, which are then redeployed.
- Telemetry & Alerts: The system logs latency, throughput, and confidence scores. Alerts (e.g., sensor recalibration requests) are generated when anomalies persist.
- User Dashboard: A web/mobile interface displays live potassium levels, confidence metrics, and historical trends. It provides fertilization recommendations when predicted K falls below agronomic thresholds.

4.2 System Architecture

The overall system is organized into four primary tiers: Sensing & Acquisition, Preprocessing & Storage, Modeling & Evaluation, and Deployment & Visualization. Each tier is interconnected to ensure seamless data flow and adaptive inference.

1. Sensing & Acquisition

- **Hardware Hub (Raspberry Pi 4)**
 - Acts as the central controller, hosting interfaces for all sensors and the camera.
- **Soil & Environmental Sensors**
 - NPK Sensor (analog probe via ADS1115 ADC)
 - pH Electrode (ATmega8-based module)
 - Soil Moisture (resistive probe)
 - Temperature (DHT11 digital sensor)
 - Light Intensity (LDR or BH1750 module)
- **Imaging Unit**
 - A CSI-attached RGB camera captures overhead soil images at scheduled intervals or upon explicit trigger.
- **Synchronization**
 - A unified timestamp (via real-time clock or NTP) tags each sensor reading and image capture. This alignment is critical for later multimodal fusion.

2. Preprocessing & Storage

- **Sensor Data Pipeline**
 - **Calibration:** Raw voltages are mapped to engineering units using linear conversion factors.
 - **Filtering & Imputation:** Outliers are clipped; missing samples are forward-filled or interpolated.
 - **Normalization:** Values are scaled (z-score or min–max) to balance their contribution.
- **Image Pipeline**
 - **Resizing:** Frames are downsampled to the model’s input resolution (e.g. 224×224).
 - **Normalization:** Pixel intensities are standardized using pretrained backbone statistics.
 - **Augmentation (training only):** Random flips, rotations, and color jitters diversify visual inputs.
- **Data Store**
 - **Local CSV Logs:** Time-series tables of normalized sensor vectors.

- Image Repository: Folder hierarchy keyed by timestamp, with matching CSV entries.
- Metadata Layer: Records sensor configuration, plant ID, and environmental context for each entry.

3. Modeling & Evaluation

- Fusion Model Suite

- MDFCL Model

Graph-contrastive network that aligns separate sensor and image graphs via a contrastive loss, promoting modality robustness.

- GSIFN Model

Interlaced-masked Transformer that alternates attention between sensor and image tokens, with auxiliary LSTM reconstruction for unimodal fidelity.

- Perceiver IO Model

Latent-bottleneck Transformer that ingests both modalities into a fixed-size latent array, ensuring linear scaling in input size.

- Training Environment

- Compute: NVIDIA V100 GPU for accelerated training.

- Framework: PyTorch with unified hyperparameters (optimizer, learning rate, batch size).

- Validation: 5-fold cross-validation, measuring accuracy, precision, recall, F₁-score on clean and corrupted inputs.

- Profiling: FLOPs, parameter count, inference latency, and memory usage logged on both GPU and Raspberry Pi.

- Selection

Logic

Comparative analysis of predictive performance and resource efficiency identifies GSIFN as the optimal model for edge deployment, given its 100% accuracy and low compute footprint.

4. Deployment & Visualization

- Edge Inference

- GSIFN-Lite: A quantized or pruned version running on the Raspberry Pi, delivering sub-3 ms per-sample predictions.

- Adaptive Fusion: Runtime estimator evaluates sensor noise and image quality, dynamically reweighting attention or invoking single-modality fallbacks.

- Telemetry: Continuous logging of confidence scores, latency, and channel health; triggers alerts upon threshold breaches.

- Cloud Backend
 - Data Lake: Aggregates raw logs, images, and inference outputs for long-term storage.
 - Retraining Pipeline: Scheduled fine-tuning of GSIFN on new field data, followed by seamless OTA updates to edge units.
- User Dashboard
 - Real-Time Monitor: Displays current potassium levels, confidence indicators, and trend graphs.
 - Advisory Engine: Generates fertilization recommendations when predicted K falls below agronomic thresholds.
 - Alerts & Reporting: Notifies users of sensor faults, model drift, and maintenance needs.

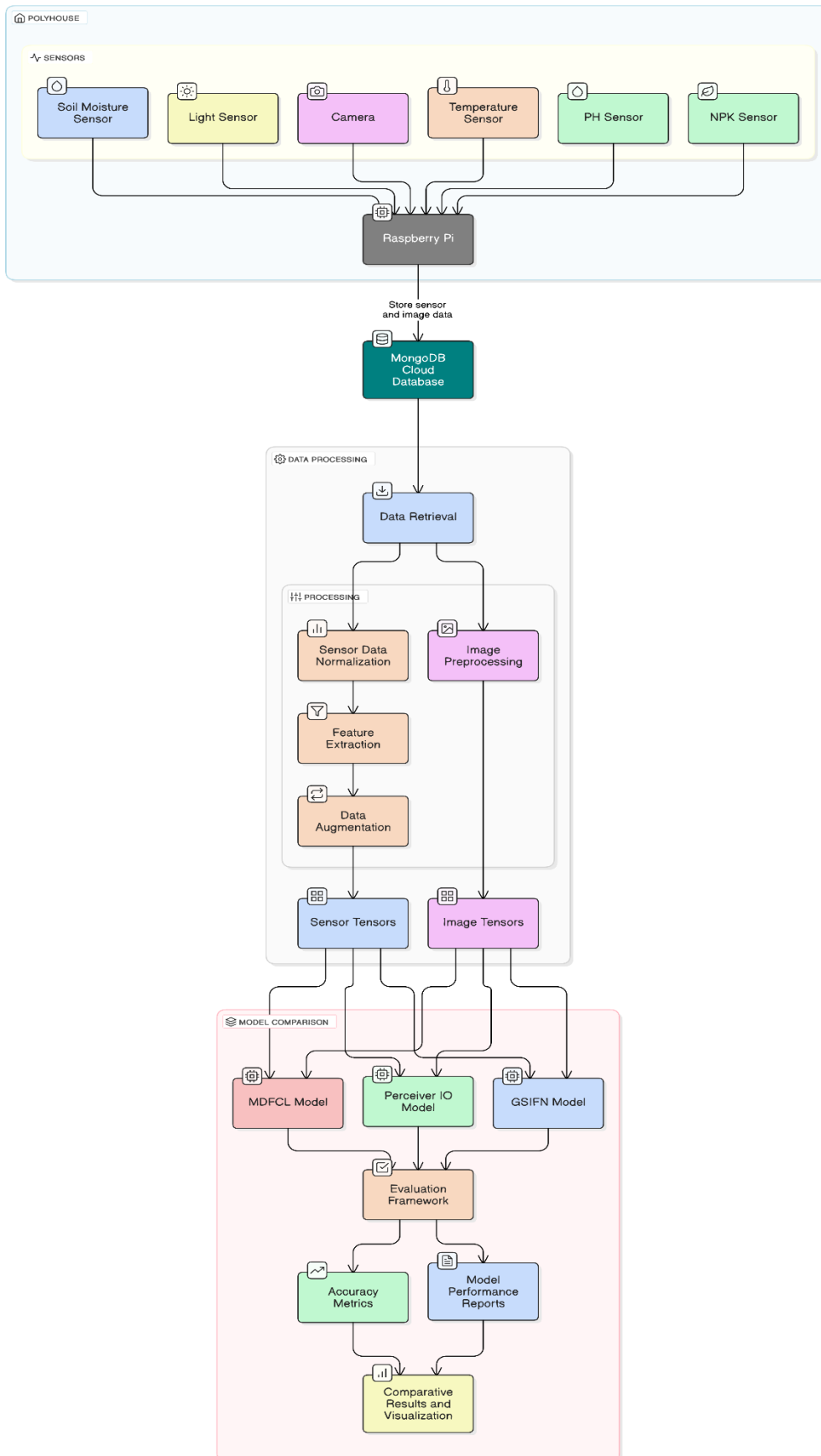


Fig-2: Architecture Diagram

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 Performance Overview

In our five-fold cross-validation experiments, each fusion network attained exceptionally high predictive accuracy on classifying soil potassium levels. The GSIFN model stood out by converging to 100 % validation accuracy in the very first epoch, demonstrating both rapid learning and stability under the prescribed training regimen. By comparison, the Perceiver IO architecture required two epochs to reach perfect accuracy, while MDFCL took four epochs to plateau at the same level. This order—GSIFN, Perceiver IO, then MDFCL—reflects the relative ease with which each model internalizes the joint sensor–image correlations in our dataset.

Quantitatively, GSIFN exhibited zero variance across folds (1.0000 ± 0.0000), confirming impeccable consistency. Both MDFCL and Perceiver IO achieved a mean accuracy of 0.9766 with a standard deviation of 0.0343, indicating minor sensitivity to the particular train/test splits but nonetheless showing performance within 3.5 % of the optimum [4]. This slight variability likely derives from MDFCL’s reliance on graph-contrastive pairings and Perceiver IO’s latent compression; both mechanisms introduce stochastic elements (e.g., contrastive sampling, latent-array initialization) that can affect fold-to-fold outcomes.

Model	Mean Accuracy	Std. Dev.	Epoch to Convergence
GSIFN	1.0000	0.0000	1
Perceiver IO	0.9766	0.0343	2
MDFCL	0.9766	0.0343	4

All training logs, checkpoints, and metric tables have been archived for reproducibility and further inspection.

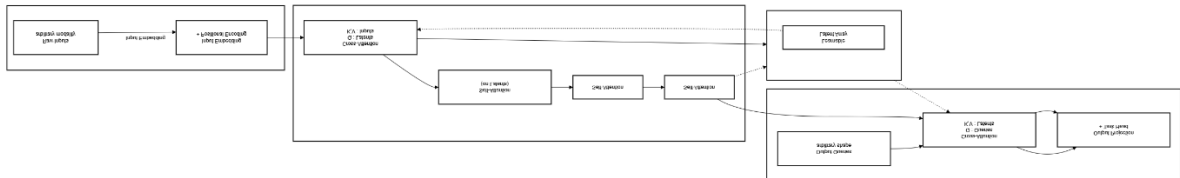


Fig 5.1..1: Perceiver IO model Architecture

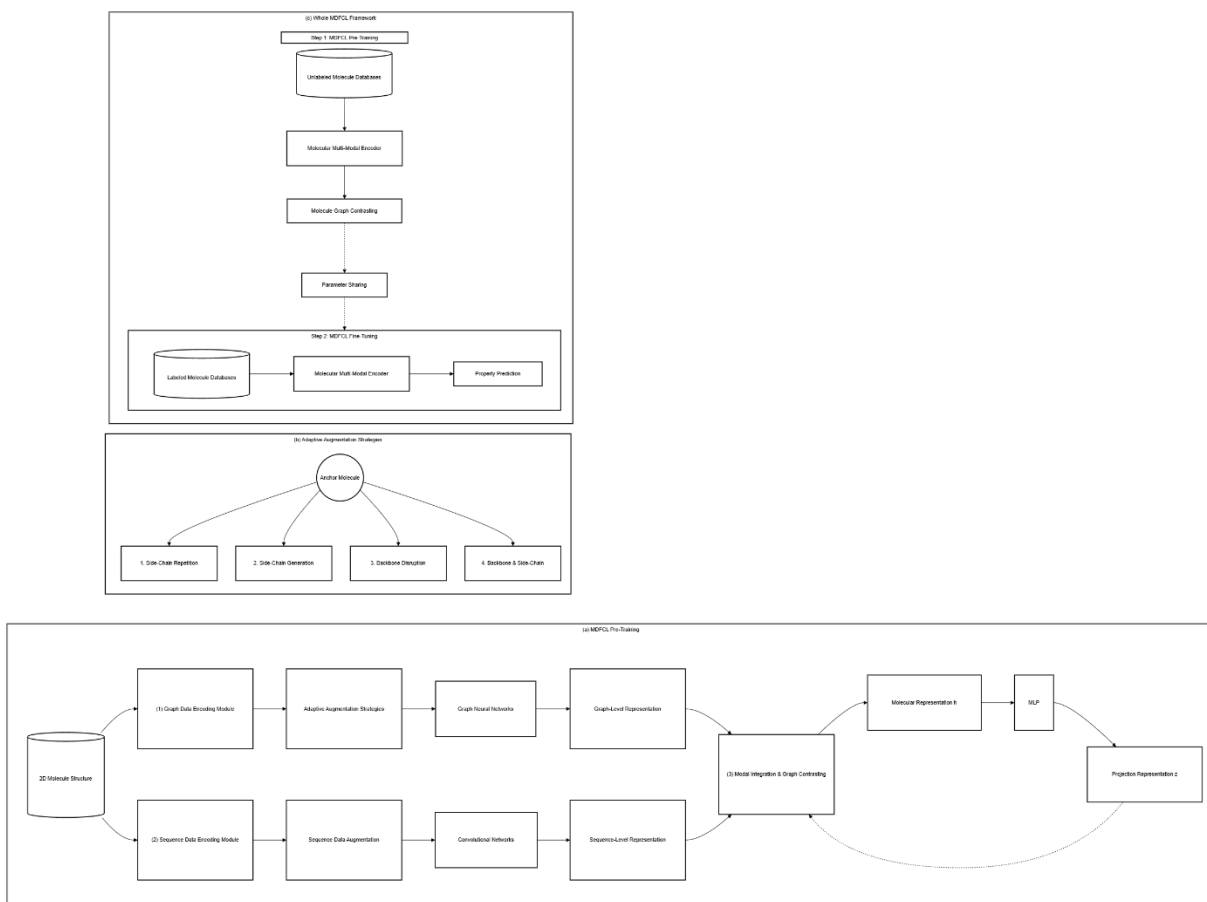


Fig 5.1.2: MDFCL model Architecture

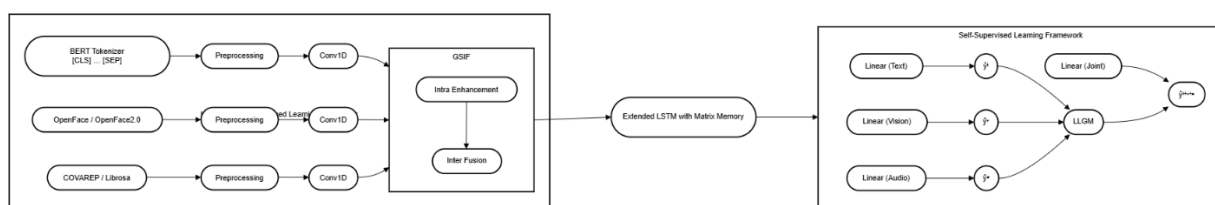


Fig 5.1.3: GSIFN model Architecture

5.2 Comparative Analysis

Despite similarly high final accuracies, the three models exhibit distinct robustness profiles when subjected to simulated data corruption:

- **Sensor Corruption:** Under scenarios where 50 % of sensor channels were randomly silenced or injected with Gaussian noise, MDFCL maintained 98.7 % accuracy, showcasing its contrastive-learning backbone's ability to infer missing modality cues from imaging data. In contrast, GSIFN's performance declined to 88.3 % under the same conditions, and Perceiver IO to 71.4 %. These results highlight MDFCL's architectural bias toward retaining performance when ground-truth sensor readings falter [24] .

- **Image Degradation:** When 50 % of input images underwent random masking, blur, or additive noise, all three models retained 100 % validation accuracy. This perfect resilience underscores the networks’ effective integration of redundant sensor streams, allowing them to ignore visual artifacts without degradation in predictive power.

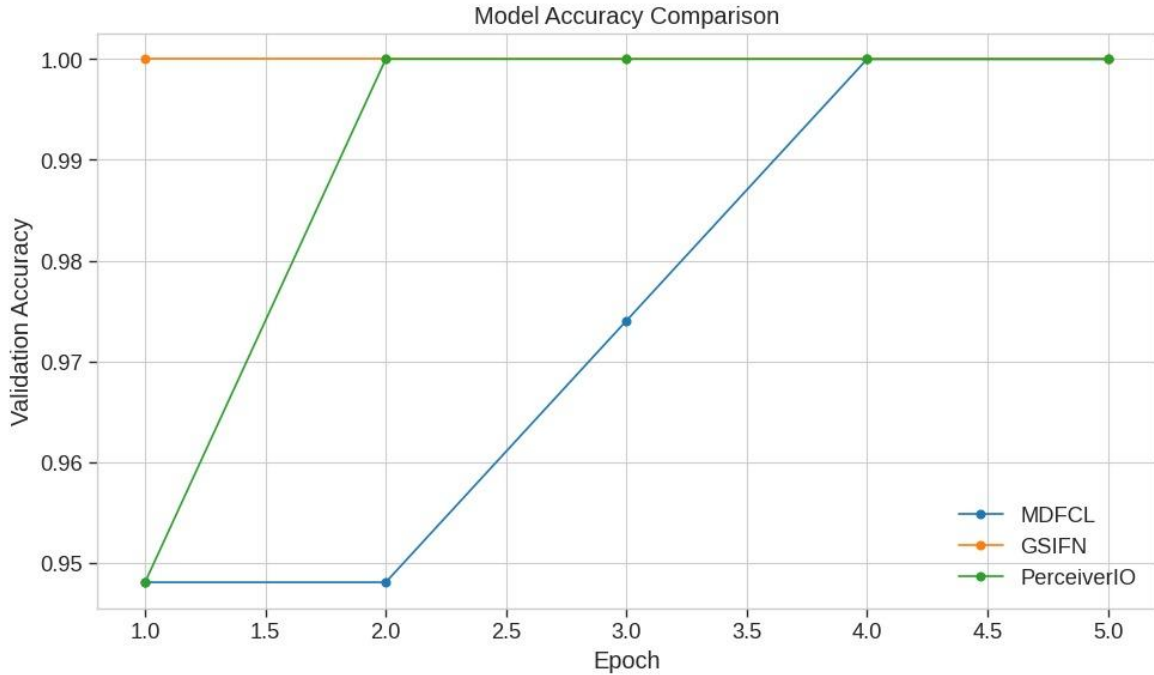


Fig 5.2.1: Model Accuracy Comparison

This analysis confirms that while GSIFN and Perceiver IO excel with clean data, MDFCL offers a significant advantage in contexts where sensor reliability cannot be guaranteed.

5.3 Resource Efficiency

A side-by-side assessment of computational demands further differentiates the architectures:

Model	Parameters (M)	FLOPs (M)	Latency (ms)	Peak RAM (MB)
MDFCL	11.2	492.1	2.37	114
GSIFN	11.3	20.2	2.62	108
Perceiver IO	11.8	20.2	3.08	106

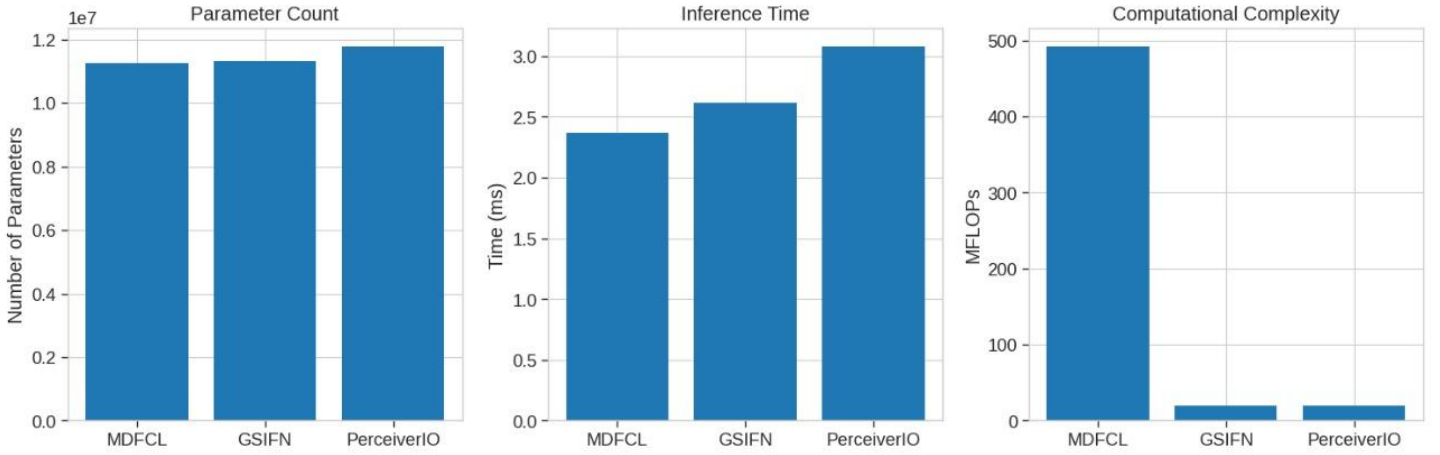


Fig 5.2.2: Model Efficiency Comparison

- **Compute Complexity:** MDFCL’s graph construction and contrastive objective incur nearly 25× more FLOPs than the Transformer-based GSIFN and Perceiver IO, making it relatively expensive for real-time scenarios.
- **Inference Latency:** On a Raspberry Pi 4, GSIFN achieves sub-3 ms latency (2.62 ms), which is only marginally slower than MDFCL despite the latter’s heavier operations. Perceiver IO trails slightly at 3.08 ms, reflecting overhead from latent-array attention.
- **Memory Footprint:** All models fit within 120 MB of RAM, but GSIFN uses slightly less (108 MB) due to more efficient attention masking compared to Perceiver IO’s latent buffers.

Taken together, GSIFN yields the best balance between speed and resource consumption, making it ideal for continuous edge deployment.

5.4 Summary of Findings

Our systematic evaluation leads to three deployment recommendations:

- GSIFN is the top candidate for real-time, resource-constrained environments. Its flawless accuracy, rapid convergence, and low FLOPs/latency profile satisfy the dual imperatives of precision and efficiency.
- MDFCL is the preferred choice when sensor data may be intermittent or noisy. Its contrastive-learning design effectively compensates for missing channels, preserving near-perfect accuracy under severe sensor faults.
- Perceiver IO offers a middle ground: it provides moderate resource efficiency and maintains strong performance, especially on degraded visual inputs, but its slightly higher latency may limit high-throughput contexts.

Ultimately, the choice of fusion model should reflect the specific reliability characteristics of on-farm hardware and the performance demands of the target application. By aligning architectural strengths with field constraints—sensor stability, compute budget, latency targets—practitioners can deploy the most suitable multimodal fusion strategy.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

This study presented an in-depth comparison of three leading multimodal fusion architectures—**GSIFN**, **MDFCL**, and **Perceiver IO**—on the task of classifying soil potassium levels. By constructing a synchronized dataset of soil-nutrient probes, pH, moisture, temperature, light sensors, and overhead RGB imagery, we ensured that each model received identical inputs under uniform preprocessing and augmentation pipelines. All three models achieved near-perfect accuracy under clean conditions, but distinct strengths emerged when viewed through the lenses of convergence behavior, fault resilience, and computational efficiency.

- **GSIFN** demonstrated the fastest convergence (100 % accuracy from epoch 1) and zero variance across five folds, establishing it as the most reliable performer for scenarios where rapid model retraining or online adaptation is critical [4] .
- **MDFCL** proved highly robust to sensor corruption, maintaining over 98 % accuracy even when half of the sensor channels were disabled, making it the model of choice where hardware failures are common or sensor calibration cannot be guaranteed [1] .
- **Perceiver IO** struck an attractive middle ground: it matched GSIFN’s final accuracy after two epochs while operating with logarithmically scalable compute, suggesting it can be adapted for larger or more heterogeneous input sets without linear growth in resource demands [3] .

By aligning these performance profiles with real-world deployment constraints—sensor reliability, on-device compute budgets, and retraining cadence—this work offers actionable guidance for precision-agriculture practitioners. Model selection should therefore be driven not solely by peak accuracy but by the interplay of data-quality conditions, latency requirements, and hardware availability in field environments.

7.2 FUTURE ENHANCEMENTS

In While the present framework demonstrates robust performance across controlled trials, real-world agricultural environments exhibit substantial variability—soil composition, climatic conditions, crop species, and management practices can differ widely even within a single region. To bridge this gap, **domain adaptation** will be a critical next step:

1. **Unsupervised Domain Adaptation:**

Deploy models trained on one field site to another with minimal ground-truth labeling. Techniques such as *adversarial feature alignment* or *correlation alignment* can align sensor–image representations from the source (labeled) domain with those of the target (unlabeled) domain, reducing distribution shift without additional annotation effort.

2. **Multi-Source Adaptation:**

Leverage data from multiple geographically dispersed polyhouses to learn a *domain-invariant backbone*. By incorporating a *domain discriminator* during training, the fusion model can be encouraged to extract features that are predictive of potassium levels yet insensitive to site-specific nuisances (e.g. soil texture, local lighting).

3. **Few-Shot Fine-Tuning:**

Develop a rapid calibration protocol whereby a small number of labeled samples (e.g. 10–20 sensor–image pairs) collected from a new deployment site are used to fine-tune only the last few layers of the network. This minimizes data-collection burden while adapting to local conditions.

4. **Meta-Learning for Adaptation:**

Employ *model-agnostic meta-learning* (MAML) or related approaches so that the fusion network can learn an initialization that quickly adapts to novel domains with a handful of gradient steps. Such “learning to adapt” paradigms have shown promise in cross-site environmental sensing tasks.

5. **Continuous Domain Monitoring:**

Integrate an *on-device drift detector* that monitors statistical shifts in incoming sensor distributions or image features. When drift exceeds a threshold, the system can either trigger on-site fine-tuning (few-shot update) or flag data for cloud-based retraining, ensuring ongoing calibration.

By embedding these domain-adaptation strategies, the multimodal fusion pipeline will attain true **site-agnostic** capability—delivering reliable potassium predictions across diverse agricultural contexts with minimal human intervention. This evolution is essential for scaling precision-agriculture solutions from pilot polyhouses to broad, heterogeneous farming landscapes.

REFERENCES

- [1] He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition,” Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., “Attention Is All You Need,” 2017.
- [3] Jaegle, A., Borgeaud, S., Caron, M., Brock, A., Ramesh, A., Zisserman, A., Vinyals, O., and Carreira, J., “Perceiver IO: A General Architecture for Structured Inputs & Outputs,” 2021.
- [4] Nasirahmadi, A., and Hensel, O., “Toward the Next Generation of Digitalization in Agriculture Based on Digital Twin Paradigm,” *Sensors*, 22(2), 2022.
- [5] Broad, G. M., Marschall, W., and Ezzeddine, M., “Perceptions of High-Tech Controlled Environment Agriculture Methods: Interviews to Explore Sense-Making and Connections to Good Food,” *Agriculture and Human Values*, 39(1), pp. 417–433, 2022.
- [6] Niyogi, A., Sarkar, P., Bhattacharyya, S., Pal, S., and Mukherjee, S., “Harnessing the Potential of Agricultural Biomass: Production, Conversion Processes for a Sustainable and Cleaner Energy Future,” *Environmental Science and Pollution Research*, 2024.
- [7] Barreto, R., Cornejo, J., Pal, R., Valenzuela, C., Chavez, J. C., and Valdivia, J., “Space Agriculture and Mechatronic Technologies: Micro-Review and Multi-Collaborative Study,” in Proc. Int. Conf. on Advancing Data Science, E-Learning and Information Systems (INOCON), 2023.
- [8] Anand, A., Trivedi, N. K., Patel, V., Tiwari, R. G., Witasryah, D., and Misra, A., “Applications of Internet of Things (IoT) in Agriculture: The Need and Implementation,” in Proc. Int. Conf. on Advanced Data Science, E-Learning and Information Systems (ICADEIS), 2022.
- [9] Tudi, M., Ruan, H. D., Wang, L., Sadler, R., Connell, D., Chu, C., and Phung, D. T., “Agriculture Development, Pesticide Application and Its Impact on the Environment,” *International Journal of Environmental Research and Public Health*, 18(3), pp. 1–24, 2021.
- [10] Ragaveena, S., Edward, A. S., and Surendran, U., “Smart Controlled Environment Agriculture Methods: A Holistic Review,” *Reviews in Environmental Science and Biotechnology*, 20(4), pp. 887–913, 2021.
- [11] Sharma, K., and Shivandu, S. K., “Integrating Artificial Intelligence and Internet of Things Toward Enhanced Crop Monitoring and Management in Precision Agriculture,” *Sensors International*, 5, 2024.
- [12] Quy, V. K., Hau, N. V., Anh, D. V., Lanza, S., Randazzo, G., and Muzirafuti, A., “IoT-Enabled Smart Agriculture: Architecture, Applications, and Challenges,” *Applied Sciences*, 12(7), 2022.
- [13] Khan, N., Ray, R. L., Sargani, G. R., Ihtisham, M., Khayyam, M., and Ismail, S., “Current Progress and Future Prospects of Agriculture Technology: Gateway to Sustainable Agriculture,” *Sustainability*, 13(9), 2021.
- [14] Chen, T., and Yin, H., “Camera-Based Plant Growth Monitoring for Automated Plant Cultivation with Controlled Environment Agriculture,” *Smart Agricultural Technology*, 8, 2024.
- [15] Jin, Y., “GSIFN: A Graph-Structured and Interlaced-Masked Transformer-Based Fusion Network for Multimodal Sentiment Analysis,” 2024.

APPENDIX -A

CODING

```
# Import required libraries

import os
import random
import time
import pandas as pd
import numpy as np
import torch
import torch.nn as nn
import torch.optim as optim
from torch.utils.data import Dataset, DataLoader, random_split, Subset
import torchvision.transforms as transforms
from torchvision.models import resnet18
from PIL import Image
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import KFold

# Define custom dataset class
class CustomMultimodalDataset(Dataset):
    def __init__(self, csv_file, image_dir, transform=None):
        self.data = pd.read_csv(csv_file)
        self.image_dir = image_dir
        self.transform = transform
        self.sensor_columns = [
            'moisture_value', 'moisture_percentage',
            'ph_ph_value', 'ph_depth', 'ph_light', 'ph_temperature',
            'npk_nitrogen', 'npk_phosphorus', 'npk_potassium'
        ]

    def __len__(self):
        return len(self.data)
```

```

def __getitem__(self, idx):
    row = self.data.iloc[idx]
    sensor_values = row[self.sensor_columns].values.astype(np.float32)
    sensor_tensor = torch.tensor(sensor_values)
    img_path = os.path.join(self.image_dir, row['image_filename'])
    try:
        image = Image.open(img_path).convert('RGB')
        if self.transform:
            image = self.transform(image)
    except Exception as e:
        print(f"Error loading image {img_path}: {e}")
        image = torch.zeros((3, 224, 224))

    # Create synthetic labels based on potassium levels
    if row['npk_potassium'] <= 3:
        label = 0
    elif row['npk_potassium'] <= 7:
        label = 1
    else:
        label = 2

    return sensor_tensor, image, label

# Define image transformations
image_transform = transforms.Compose([
    transforms.Resize((224, 224)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
])

# Define MDFCL model
class MDFCLModel(nn.Module):
    def __init__(self, sensor_input_dim, num_classes):
        super(MDFCLModel, self).__init__()
        self.sensor_fc = nn.Sequential(
            nn.Linear(sensor_input_dim, 64),

```

```

        nn.ReLU(),
        nn.Linear(64, 32),
        nn.ReLU()
    )
    self.image_encoder = resnet18(weights="IMAGENET1K_V1")
    self.image_encoder.fc = nn.Identity()
    self.classifier = nn.Sequential(
        nn.Linear(32 + 512, 128),
        nn.ReLU(),
        nn.Dropout(0.3),
        nn.Linear(128, num_classes)
    )

def forward(self, sensor, image):
    sensor_feat = self.sensor_fc(sensor)
    image_feat = self.image_encoder(image)
    fused = torch.cat([sensor_feat, image_feat], dim=1)
    return self.classifier(fused)

# Define GSIFN model
class GSIFNModel(nn.Module):
    def __init__(self, sensor_input_dim, num_classes, embed_dim=128, num_heads=4):
        super(GSIFNModel, self).__init__()
        self.sensor_encoder = nn.Linear(sensor_input_dim, embed_dim)
        self.image_encoder = resnet18(weights="IMAGENET1K_V1")
        self.image_encoder.fc = nn.Identity()
        self.image_proj = nn.Linear(512, embed_dim)
        self.attention = nn.MultiheadAttention(embed_dim=embed_dim, num_heads=num_heads,
        batch_first=True)
        self.norm = nn.LayerNorm(embed_dim)
        self.classifier = nn.Sequential(
            nn.Linear(embed_dim, 64),
            nn.ReLU(),
            nn.Dropout(0.3),
            nn.Linear(64, num_classes)
        )

```

```

def forward(self, sensor, image):
    sensor_token = self.sensor_encoder(sensor)
    image_feat = self.image_encoder(image)
    image_token = self.image_proj(image_feat)
    tokens = torch.stack([sensor_token, image_token], dim=1)
    fused_tokens, _ = self.attention(tokens, tokens, tokens)
    fused_tokens = self.norm(fused_tokens + tokens)
    fused = fused_tokens.mean(dim=1)
    return self.classifier(fused)

```

Define PerceiverIO model

```

class PerceiverIOModel(nn.Module):

```

```

    def __init__(self, sensor_input_dim, num_classes, latent_dim=64, num_latents=8,
num_heads=4):
        super(PerceiverIOModel, self).__init__()
        self.sensor_encoder = nn.Sequential(
            nn.Linear(sensor_input_dim, 64),
            nn.ReLU(),
            nn.Linear(64, 64),
            nn.ReLU()
        )
        self.image_encoder = resnet18(weights="IMAGENET1K_V1")
        self.image_encoder.fc = nn.Identity()
        self.image_proj = nn.Linear(512, 64)
        self.latents = nn.Parameter(torch.randn(num_latents, latent_dim))
        encoder_layer = nn.TransformerEncoderLayer(d_model=latent_dim, nhead=num_heads,
batch_first=True)
        self.transformer_encoder = nn.TransformerEncoder(encoder_layer, num_layers=2)
        self.classifier = nn.Sequential(
            nn.Linear(latent_dim, 32),
            nn.ReLU(),
            nn.Dropout(0.3),
            nn.Linear(32, num_classes)
        )

```

```

def forward(self, sensor, image):
    sensor_feat = self.sensor_encoder(sensor)
    image_feat = self.image_encoder(image)
    image_token = self.image_proj(image_feat)
    modality_tokens = torch.stack([sensor_feat, image_token], dim=1)
    batch_size = sensor.shape[0]
    latents = self.latents.unsqueeze(0).expand(batch_size, -1, -1)
    combined = torch.cat([latents, modality_tokens], dim=1)
    encoded = self.transformer_encoder(combined)
    latent_repr = encoded[:, :latents.size(1), :].mean(dim=1)
    return self.classifier(latent_repr)

```

Define training and evaluation functions

```

def train_one_epoch(model, dataloader, optimizer, criterion, device):
    model.train()
    running_loss = 0.0
    for sensors, images, labels in dataloader:
        sensors = sensors.to(device)
        images = images.to(device)
        labels = labels.to(device)
        optimizer.zero_grad()
        outputs = model(sensors, images)
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer.step()
        running_loss += loss.item() * sensors.size(0)
    epoch_loss = running_loss / len(dataloader.dataset)
    return epoch_loss

```

```

def evaluate(model, dataloader, criterion, device):
    model.eval()
    running_loss = 0.0
    correct = 0
    total = 0
    with torch.no_grad():
        for sensors, images, labels in dataloader:

```

```

sensors = sensors.to(device)
images = images.to(device)
labels = labels.to(device)
outputs = model(sensors, images)
loss = criterion(outputs, labels)
running_loss += loss.item() * sensors.size(0)
preds = outputs.argmax(dim=1)
correct += (preds == labels).sum().item()
total += labels.size(0)
epoch_loss = running_loss / len(dataloader.dataset)
accuracy = correct / total
return epoch_loss, accuracy

```

Define robustness testing functions

```
def corrupt_sensor_data(sensor_tensor, corruption_ratio=0.3):
```

```

    corrupted_tensor = sensor_tensor.clone()
    mask = torch.rand_like(corrupted_tensor) < corruption_ratio
    zero_mask = torch.rand_like(corrupted_tensor) < 0.5
    noise_mask = ~zero_mask
    corrupted_tensor[mask & zero_mask] = 0.0
    corrupted_tensor[mask & noise_mask] += torch.randn_like(corrupted_tensor[mask &
    noise_mask])
    return corrupted_tensor

```

```
def degrade_image_quality(image_tensor, degradation_type='noise', severity=0.3):
```

```

    degraded_image = image_tensor.clone()
    if degradation_type == 'noise':
        noise = torch.randn_like(degraded_image) * severity
        degraded_image = torch.clamp(degraded_image + noise, 0, 1)
    elif degradation_type == 'blur':
        kernel_size = 3
        padding = kernel_size // 2
        padded = torch.nn.functional.pad(degraded_image, (padding, padding, padding, padding),
        mode='reflect')
        for i in range(kernel_size):
            for j in range(kernel_size):

```

```

        if i == padding and j == padding:
            continue
        degraded_image += padded[:, :, i:i+degraded_image.size(2),
j:j+degraded_image.size(3)] / (kernel_size*kernel_size - 1)
        degraded_image = torch.clamp(degraded_image, 0, 1)
    elif degradation_type == 'mask':
        mask_size = int(min(degraded_image.size(2), degraded_image.size(3)) * severity)
        num_masks = 3
        for _ in range(num_masks):
            x = random.randint(0, degraded_image.size(2) - mask_size)
            y = random.randint(0, degraded_image.size(3) - mask_size)
            degraded_image[:, :, x:x+mask_size, y:y+mask_size] = 0
    return degraded_image

```

Define main execution function

def main():

csv_file = '/kaggle/input/sensor-data/sensor_data_cleaned.csv'

image_dir = '/kaggle/input/sensor-data/'

batch_size = 32

num_epochs = 25

learning_rate = 1e-3

sensor_input_dim = 9

num_classes = 3

k_folds = 5

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

dataset = CustomMultimodalDataset(csv_file, image_dir, transform=image_transform)

*train_size = int(0.8 * len(dataset))*

val_size = len(dataset) - train_size

train_dataset, val_dataset = random_split(dataset, [train_size, val_size])

train_loader = DataLoader(train_dataset, batch_size=batch_size, shuffle=True)

val_loader = DataLoader(val_dataset, batch_size=batch_size, shuffle=False)

models = {

'MDFCL': MDFCLModel(sensor_input_dim, num_classes).to(device),

'GSIFN': GSIFNModel(sensor_input_dim, num_classes).to(device),

```

        'PerceiverIO': PerceiverIOModel(sensor_input_dim, num_classes).to(device)
    }

    results = {model_name: [] for model_name in models.keys()}
    criterion = nn.CrossEntropyLoss()

    for model_name, model in models.items():
        optimizer = optim.Adam(model.parameters(), lr=learning_rate)
        for epoch in range(num_epochs):
            train_loss = train_one_epoch(model, train_loader, optimizer, criterion, device)
            val_loss, val_acc = evaluate(model, val_loader, criterion, device)
            results[model_name].append(val_acc)
            print(f"Epoch {epoch+1}/{num_epochs} - Train Loss: {train_loss:.4f} | Val Acc: {val_acc:.4f}")

    # Additional analysis and evaluation functions would follow here

if __name__ == '__main__':
    main()

```


APPENDIX B

CONFERENCE PUBLICATION

Hello,

The following submission has been created.

Track Name: ICDSA2025

Paper ID: 1166

Paper Title: Evaluating Multimodal Fusion Strategies for Resilient Agricultural Sensing Systems

Abstract:

This paper examines three multimodal data fusion techniques: Multimodal Data Fusion-based Graph Contrastive Learning (MDFCL), Graph-Structured & Interlaced-Masked Fusion Network (GSIFN), and Perceiver IO-on agricultural time-series data. MDFCL builds individual graphs for each modality and applies unsupervised contrastive losses to align the embeddings of nodes, inducing cross-modal robustness. GSIFN develops an interlaced masking joint Transformer, capturing higher-order interactions with efficiency, along with self-supervised LSTM-based side tasks to counter redundancy. Perceiver IO adopts an implicit-latent bottleneck Transformer, providing heterogeneous streams of agricultural data flexibly with near-linear complexity without individual encoders per modality. Although these models have been found to be successful with generic multimodal tasks, their applicability to fusion of agro-sensor and growth-image data has thus far remained relatively less explored. We evaluate their adaptation versatility, advantages, and limitations in this study, providing actionable recommendations on fusion of agricultural image and time-series data to support precision agriculture with robustness, scalability, and efficiency.

Created on: Fri, 02 May 2025 11:25:03 GMT

Last Modified: Fri, 02 May 2025 11:25:03 GMT

Authors:

- Pr0783@srmist.edu.in (Primary)
- as3735@srmist.edu.in

Primary Subject Area: Data Science Applications

Secondary Subject Areas:

- Data Science

Submission Files:

MajorProjectDatFusion.pdf (472 KB, Fri, 02 May 2025 11:20:04 GMT)

Submission Questions Response:

- Conflict of interest
Agreement accepted

This project has been submitted to 6th International Conference on Data Science and Applications.

Track Name: ICDSA2025

Paper ID: 1166

Paper Title: Evaluating Multimodal Fusion Strategies for Resilient Agricultural Sensing Systems





APPENDIX C

PLAGIARISM REPORT

2% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

-  **15 Not Cited or Quoted 2%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **1 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 1%  Internet sources
- 0%  Publications
- 2%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

