Dharti Mitra: An Uncertainty-Aware Multimodal Edge–Cloud Framework for Intelligent Crop Health Monitoring and Precision Farming

Yashkumar Lembhe
Department of Information Technology
MIT School of Computing, MIT-ADT University
Pune, Maharashtra, India
yashlembhe7@gmail.com

Krishna Pawar
Department of Information Technology
MIT School of Computing, MIT-ADT University
Pune, Maharashtra, India
krishna.pawar312006@gmail.com

Samruddhi Jadhav
Department of Information Technology
MIT School of Computing, MIT-ADT University
Pune, Maharashtra, India
samruddhij0409@gmail.com

Aarti Pimpalkar
Department of Information Technology
MIT School of Computing, MIT-ADT University
Pune, Maharashtra, India
aarti.pimpalkar@mituniversity.edu.in

***Abstract*—Smallholder farmers may need to assess crop health with incomplete field observations, limited access to agronomic expertise, and intermittent network connectivity. Image-based plant-disease classifiers provide useful screening capabilities, but controlled-image performance does not necessarily transfer to heterogeneous field photographs. Moreover, a leaf image alone may not capture contextual variables such as crop variety, growth stage, soil conditions, recent weather, and previous farm observations. This paper presents *Dharti Mitra*, a reliability-centred multimodal edge–cloud framework for crop-health monitoring and precision-agriculture decision support. The proposed research methodology combines image-derived representations with available agricultural context through mask-aware feature fusion, followed by probability calibration and selective prediction. Low-quality, uncertain, or out-of-distribution inputs are deferred for image recapture or expert review. The framework also specifies a temporal disease-risk component, a constrained recommendation interface, and an edge–cloud execution model for low-connectivity settings. A source-code audit distinguishes the existing web-application prototype from research components that still require implementation. In the inspected prototype, crop-image analysis is a colour-distribution heuristic and weather values are simulated; no trained disease model or validated performance results are available. Accordingly, this manuscript contributes a proposed architecture, formal problem specification, implementation audit, and leakage-resistant evaluation protocol rather than unverified accuracy or field-impact claims. The research questions centre on missing-modality robustness, calibrated selective prediction under field-domain shift, and the operational trade-offs of edge–cloud deployment.**

***Index Terms*—precision agriculture, plant disease detection, multimodal learning, missing modalities, uncertainty estimation, selective prediction, calibration, edge computing, crop-health monitoring.**

I. Introduction

Agricultural decision-making increasingly uses digital observations from mobile devices, sensors, farm records, and remote-sensing systems. Smart farming research has identified data interoperability, information infrastructure, and decision support as central concerns in this transition [1], [2]. In crop protection, image-based deep learning has demonstrated the feasibility of classifying visible plant-disease symptoms. Mohanty *et al.* evaluated deep convolutional models using the PlantVillage collection [3], while subsequent studies compared architectures and training strategies for plant-disease recognition [4], [5].

However, performance on curated images does not guarantee reliable operation in farms. Field images may contain complex backgrounds, multiple leaves, variable illumination, occlusion, mixed symptoms, and previously unseen cultivars. PlantDoc was introduced to provide non-laboratory images for plant disease recognition [6]; FieldPlant subsequently provided plantation-collected images with leaf-level annotations supervised by plant pathologists [7]. Recent cross-domain studies have quantified large reliability degradation when models trained on controlled imagery are evaluated on field photographs [8], [9].

A second challenge is that crop-health interpretation may depend on context. Crop identity, growth stage, soil conditions, weather, irrigation, and past observations can inform the interpretation of visual symptoms. At the same time, such data are often incomplete: a small farm may not have soil sensors, weather observations may be delayed, and connectivity may be unavailable. Recent studies have already explored multimodal disease diagnosis using agro-ecological metadata, visual–text information, and agrometeorological variables [10]–[13]. Consequently, multimodal fusion alone is not claimed as the novelty of this work.

This paper positions *Dharti Mitra* as a reliability-centred system study. Its central question is whether a practical crop-health platform can jointly handle partially available modalities,

field-domain uncertainty, conservative decision deferral, and intermittent edge–cloud connectivity. The proposed method combines modality-specific encoders with availability-aware fusion, uses a separately fitted calibration stage, and applies explicit acceptance conditions before an automated prediction is shown. A rule-constrained recommendation layer is designed to present only agronomically reviewed options; it is not presented as a causal treatment optimizer.

A. Problem Statement

Given a crop image and a variable set of contextual observations, the system must estimate a crop-health class while accounting for image quality, missing inputs, and distribution shift between development and deployment data. It must also determine when the evidence is insufficient for automated guidance. In such cases, the system should request a better image or route the case to an agricultural expert instead of forcing a class prediction. The system must support a local image-inference path and synchronize records when connectivity returns.

B. Research Questions

The study is guided by the following questions:

RQ1. Does context-aware multimodal fusion improve field-image classification over an image-only baseline when the contextual labels are available and aligned?

RQ2. How does classification and calibration performance change as weather, soil, or historical modalities are missing or corrupted?

RQ3. Can a selective decision policy reduce error among accepted predictions at a useful prediction coverage under field-domain shift?

RQ4. What are the latency, memory, model-size, and availability trade-offs between local image inference and cloud-assisted contextual analysis?

RQ5. Can the recommendation layer remain within a domain-expert-approved action set and communicate uncertainty without presenting association as causation?

C. Research Contributions and Scope

The proposed contributions are deliberately framed as a system and evaluation agenda rather than as claims of a new CNN, calibration method, or general multimodal learning algorithm:

1) a reference architecture connecting visual observations and optional agricultural context using an explicit availability mask;
2) a reliability pipeline that separates image-quality rejection, probability calibration, out-of-distribution assessment, and selective prediction;
3) an evaluation protocol that measures performance under farm-/plant-aware splits, controlled-to-field transfer, and systematic modality removal;
4) an edge–cloud execution design that specifies local inference, queued synchronization, model-version tracking, and degraded-connectivity behaviour;
5) a source-code-grounded implementation audit that separates verified prototype functions from research extensions requiring implementation and validation.

The novelty of the combined system and protocol must be confirmed through a systematic search and comparative experiments. The individual building blocks are established in prior work; the research value rests on their coordinated evaluation under the stated operational constraints.

II. Related Work

A. Image-Based Disease Recognition and Field Datasets

Mohanty *et al.* demonstrated deep-learning-based plant-disease recognition with PlantVillage images [3]. Ferentinos [4] and Too *et al.* [5] examined deep architectures and fine-tuning strategies. Such work establishes image classification as a useful baseline, but data acquisition conditions must be considered when interpreting performance. PlantDoc contains 2,598 images across 13 plant species and up to 17 disease classes, with annotations on images collected from non-laboratory sources [6]. FieldPlant contains 5,170 images collected directly from plantations and 8,629 manually annotated leaves across 27 disease classes [7]. These datasets motivate external field testing and careful label harmonization.

B. Multimodal Agricultural Diagnosis

Recent literature makes it necessary to narrow the claim for multimodal novelty. Zakeeya and Sugali describe ContextLeafXAI, which combines leaf imagery with weather, soil, and geographic context using reliability-aware gated fusion and explainability methods [10]. Wang and Liu introduce AgriMM, integrating visual information with expert-validated textual descriptions [11]. Odounfa and Glèlè Kakaí report a multi-task chili-pepper framework that combines images and agrometeorological information for disease diagnosis and severity estimation [12]. Manda *et al.* review multimodal learning and foundation models for plant disease detection and forecasting [13]. These studies show that image–context fusion and multimodal disease modelling are active research directions. *Dharti Mitra* therefore does not claim to introduce multimodal learning itself. Instead, it proposes to evaluate the joint effects of partial modality availability, conservative abstention, field-domain testing, and edge–cloud operation within a farmer workflow.

C. Uncertainty, Calibration, and Domain Shift

Temperature scaling is an established post-hoc calibration method for neural classifiers [14]. Selective classification formalizes the option to abstain when estimated risk is too high [15]. Ovadia *et al.* studied predictive uncertainty under dataset shift [16]. In the agricultural domain, Shrivastava *et al.* evaluate a hierarchical vision transformer with Monte Carlo dropout and calibration across controlled and field datasets [9]. Xiang *et al.* benchmark calibration, selective prediction, and other reliability methods under controlled-to-field shift and report that standard post-hoc techniques may not close the gap [8]. This evidence cautions against treating a high softmax

score or a calibrated source-domain probability as sufficient assurance for field deployment.

D. Smart Farming and Edge–Cloud Systems

Smart-farming platforms integrate data acquisition, management systems, and decision support across farm operations [1]. In a resource-constrained deployment, edge inference can reduce dependence on continuous connectivity, while cloud services can support more expensive historical analysis and model management. The present work specifies this as an architectural requirement; the inspected prototype does not yet contain a validated mobile inference model. Federated learning is considered only as a possible future extension and is not claimed as an implemented capability.

E. Research Gap and Positioning

Table I compares representative studies against the concerns addressed here. The literature already contains image–metadata fusion, image–text fusion, uncertainty estimation, and cross-domain reliability analysis. A defensible gap is therefore narrower: a common, reproducible system-level evaluation of (i) missing contextual modalities, (ii) quality- and OOD-aware selective deferral, (iii) separate calibration and external field testing, and (iv) edge–cloud execution constraints. Whether this combination provides measurable benefits remains an empirical question.

III. Problem Formulation

Let the ith farm observation be

$$X_i = (I_i, M_i, W_i, S_i, H_i), \tag{1}$$

where I_i is an RGB crop image, M_i contains crop identity, variety and growth stage, W_i contains time-aligned weather observations, S_i contains soil or irrigation observations, and H_i contains time-stamped historical crop-health records. The disease label $Y_i \in \{1, \ldots, K\}$ must be verified by an appropriately qualified annotator. The image is required for image-based diagnosis after passing a quality gate; contextual modalities may be unavailable. Let

$$\mathbf{m}_i = (1, m_{i,M}, m_{i,W}, m_{i,S}, m_{i,H}) \in \{0, 1\}^5 \tag{2}$$

be the modality-availability vector, where the first component denotes the required image modality. The primary prediction task is

$$\hat{\mathbf{p}}_i = P(Y_i \mid X_i, \mathbf{m}_i), \tag{3}$$

subject to controlled false-acceptance risk. Secondary tasks are (i) a short-horizon disease-event forecast from time-aligned environmental and historical data, and (ii) ranking of agronomically admissible actions. The framework must degrade gracefully when contextual inputs or cloud connectivity are absent.

IV. Proposed Methodology

A. Data Quality, Provenance, and Preprocessing

Each image should retain capture time, farm/plant identifier, crop identity, source device, and annotation provenance. A quality module checks minimum resolution, blur, exposure, occlusion, and whether a relevant leaf region is visible. The quality model must be trained or calibrated on independently labelled usable/unusable images; a hand-coded threshold may be used only as a prototype gate and must be described as such. Images failing the gate are returned for recapture rather than assigned a disease label.

For training, resizing and normalization must follow the selected backbone. Augmentation may include moderate geometry and photometric transformations, but transformations must not erase or invent disease symptoms. Structured features are normalized using training-partition statistics only. Timestamps must be aligned to crop observations, and simulated weather must be stored with an explicit simulation flag rather than merged silently with measured observations. Missingness indicators are retained as model inputs.

B. Visual and Contextual Encoders

A visual encoder maps the accepted image to a latent vector:

$$\mathbf{z}_I = f_\vartheta(I) \in \mathrm{R}^{d_I}. \tag{4}$$

Candidate image backbones include ResNet-50, MobileNetV3, EfficientNet-B0, and a compact hierarchical transformer. Selection must be based on grouped field validation and device cost, not on controlled-image accuracy alone. Separate encoders transform structured metadata, weather sequences, soil measurements, and history:

$$\mathbf{z}_M = g_\phi(M) \in \mathrm{R}^{d_M}, \quad \mathbf{z}_W = h_\psi(W) \in \mathrm{R}^{d_W},$$
$$\mathbf{z}_S = q_\xi(S) \in \mathrm{R}^{d_S}, \quad \mathbf{z}_H = r_\omega(H) \in \mathrm{R}^{d_H}. \tag{5}$$

The structured encoders should be selected to match data type: categorical embeddings and an MLP for crop metadata, a temporal encoder for weather sequences, and a feature encoder for soil and historical observations.

Each available representation is projected into a common d-dimensional space:

$$\tilde{\mathbf{z}}_j = \mathbf{P}_j \mathbf{z}_j + \mathbf{b}_j, \quad j \in \{I, M, W, S, H\}. \tag{6}$$

The matrices $\mathbf{P}_j$ must map from the encoder-specific dimension to R^d.

C. Availability-Aware Feature Fusion

A trainable compatibility score is assigned to each projected embedding:

$$e_{ij} = \mathbf{u}^\top \tanh(\mathbf{A}\tilde{\mathbf{z}}_{ij} + \mathbf{b}_a), \tag{7}$$

where $\mathbf{u}$, $\mathbf{A}$, and $\mathbf{b}_a$ are learned parameters. The availability-aware weight is

$$\alpha_{ij} = \frac{m_{ij} \exp(e_{ij})}{\sum_{k \in \mathrm{J}} m_{ik} \exp(e_{ik})}, \tag{8}$$

TABLE I
REPRESENTATIVE RESEARCH RELEVANT TO DHARTI MITRA AND ITS REMAINING EVALUATION QUESTIONS.

Study	Primary focus	Data / method	Relevance and remaining question for this work
Mohanty *et al.* [3]	Image disease classification	PlantVillage; deep CNNs	Foundational image-only baseline; controlled-image results require external field evaluation.
Singh *et al.* [6]	Field-image dataset	PlantDoc; 2,598 images	Supports in-the-wild evaluation; label harmonization and grouped splits remain important.
Moupojou *et al.* [7]	Field-image dataset	FieldPlant; 5,170 images and 8,629 leaf annotations	Supports expert-supervised field benchmarking.
Zakeeya and Sugali [10]	Context-aware multimodal diagnosis	Leaf images plus weather, soil, and geographic metadata	Demonstrates that agro-ecological fusion is prior art; motivates a sharper focus on missing-data and deployment evaluation.
Wang and Liu [11]	Visual–text diagnosis	Images and expert-validated descriptions	Establishes visual–text multimodality; does not eliminate the need to test sensor/weather availability.
Odounfa and Glèlè Kakaí [12]	Diagnosis and severity estimation	Chili images plus agrometeorological data	Relevant multi-task comparator; crop-specific external validation remains necessary.
Shrivastava *et al.* [9]	Cross-domain uncertainty	Hierarchical ViT, MC dropout, calibration	Shows uncertainty-aware field evaluation is active prior work.
Xiang *et al.* [8]	Reliability under controlled-to-field shift	Calibration, selective prediction, OOD/adaptation benchmarks	Shows standard mitigation may be insufficient; motivates operational risk and coverage measurement.
Dharti Mitra (proposed)	System-level reliability evaluation	Masked multimodal inputs, selective deferral, edge–cloud protocol	To be validated through grouped field tests, missing-modality ablations, and device/network benchmarks; no result is claimed here.

TABLE II
TARGETED RESEARCH QUESTIONS AND VALIDATION PLAN. THESE ARE EVALUATION TARGETS, NOT ESTABLISHED GAPS OR COMPLETED RESULTS.

Evaluation axis	What prior work establishes	Research question	Planned validation
Partial context availability	Multimodal image–metadata and image–text systems already exist [10]–[12].	RQ2	Train with controlled modality dropout; test naturally missing and synthetically masked weather, soil, and history inputs.
Field-domain reliability	Cross-domain uncertainty and reliability benchmarking have been reported [8], [9].	RQ1, RQ3	Use grouped external field tests; report calibration, OOD metrics, coverage–risk curves, and error analysis.
Operational connectivity	Edge inference is a known deployment strategy; this prototype has no validated mobile model.	RQ4	Benchmark local image inference, queued synchronization, interrupted requests, latency, RAM, and offline success on named devices.
Action safety	Crop recommendations depend on local agronomy and product-specific restrictions.	RQ5	Constrain outputs to an expert-approved action catalogue; audit unsupported actions and escalation decisions.

where $\mathcal{J} = \{I, M, W, S, H\}$ and $m_{il} = 1$ after the image-quality gate. The fused representation is

$$\mathbf{z}_{F,i} = \sum_{j \in \mathcal{J}} \alpha_{ij} \tilde{\mathbf{z}}_{ij}. \tag{9}$$

A contextual modality with $m_{ij} = 0$ receives zero weight. Because the image is mandatory for image-based disease classification, the denominator in (8) remains non-zero after the quality gate. If all contextual modalities are missing, (9) reduces to the image-only representation. This fallback must be explicitly tested.

To expose the model to partial observations during learning, training may sample contextual availability masks from a predeclared missingness curriculum. A consistency term can be studied as an ablation:

$$L_{\text{cons}} = D_{\text{KL}}\left(\mathbf{p}_i^{\text{full}} \,\|\, \mathbf{p}_i^{\text{masked}}\right), \tag{10}$$

where the full-context and masked-context predictions are generated for the same training sample. This term is a proposed experiment, not a validated improvement; its weight must be selected on training/validation data and its risk of suppressing useful context must be examined.

D. Classification Objective

For N labelled samples and class weight w_{y_i}, the weighted cross-entropy objective is

$$L_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^{N} w_{y_i} \log p_{i,y_i}. \tag{11}$$

If severity and image-usability labels are independently annotated, auxiliary losses may be added:

$$L_{\text{total}} = L_{\text{cls}} + \lambda_s L_{\text{sev}} + \lambda_q L_{\text{qual}} + \lambda_c L_{\text{cons}}. \tag{12}$$

Terms without corresponding labels or model heads must be omitted. Hyper- parameters are selected using validation data only. Report the exact loss components actually used in any future implementation.

E. Calibration and Selective Prediction

Let $\mathbf{z}_i$ denote classifier logits. Temperature scaling produces

$$p_T(y = k \mid x_i) = \frac{\exp(z_{ik}/T)}{\sum_{j=1}^{K} \exp(z_{ij}/T)}, \tag{13}$$

where $T > 0$ is fitted by minimizing negative log-likelihood on a calibration partition that is disjoint from model training and

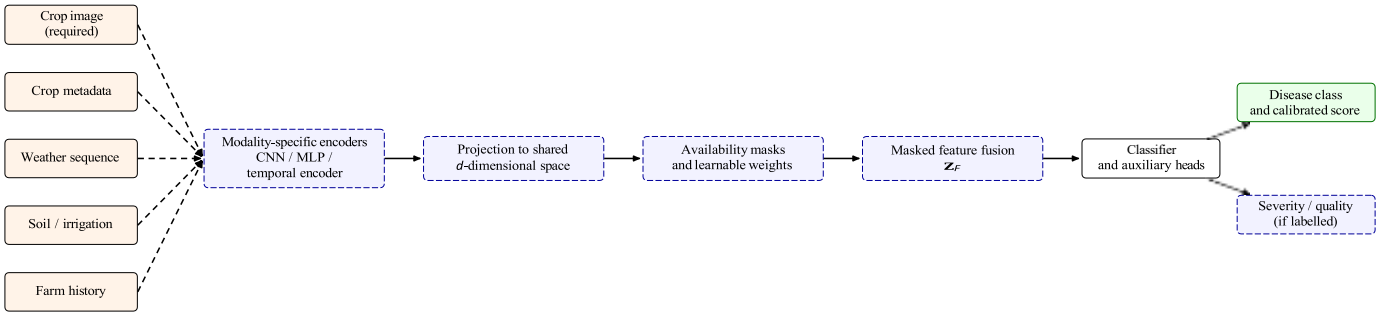


Fig. 1. Proposed availability-aware multimodal classifier. The crop image is required for image-based diagnosis; crop metadata, weather, soil, and history are optional contextual modalities. Dashed blue boxes denote research components requiring implementation and validation.

final testing [14]. Expected calibration error (ECE) is estimated as

$$\mathrm{ECE} = \sum_{b=1}^{B} \frac{|S_b|}{N} \left|\mathrm{acc}(S_b) - \mathrm{conf}(S_b)\right|, \tag{14}$$

where S_b is the set of predictions in confidence bin b. ECE should be reported with the binning scheme because it is sensitive to bin choice; the multiclass Brier score and negative log-likelihood should also be reported.

Let $q(x)$ be an image-quality score, $c(x) = \max_k p_T(y = k \mid x)$ be the maximum calibrated class probability, and $d(x)$ be a normalized OOD score whose direction is defined so that larger values indicate greater novelty. Rather than treating their product as a probability, the system uses an explicit three-condition acceptance gate:

$$g_\tau(x) = \mathbb{1}[q(x) \geq \tau_q]\, \mathbb{1}[c(x) \geq \tau_c]\, \mathbb{1}[d(x) \leq \tau_d]. \tag{15}$$

Thresholds are selected on a development/calibration cohort under a predeclared risk–coverage objective. For a test set of size N, coverage and selective risk are

$$\mathrm{Coverage}(\tau) = \frac{1}{N} \sum_{i=1}^{N} g_\tau(x_i), \tag{16}$$

$$\mathrm{Risk}(\tau) = \frac{\sum_{i=1}^{N} g_\tau(x_i)\, \mathbb{1}[\hat{y}_i \neq y_i]}{\sum_{i=1}^{N} g_\tau(x_i)}, \tag{17}$$

provided at least one prediction is accepted. Report coverage and risk jointly; accepted accuracy alone can hide excessive abstention. Calibration on a controlled validation set must not be assumed to remain valid on field data.

F. Temporal Disease-Risk Forecasting

Let $\mathbf{v}_{t-L+1:t}$ be a time-aligned sequence of weather and crop-state features over a look-back window of length L. A temporal model estimates the probability of a defined future disease event:

$$\hat{r}_{t+h} = \mathrm{sigmoid}(F_\eta(\mathbf{v}_{t-L+1:t})), \tag{18}$$

where h is the forecast horizon. The binary target must be defined as an expert-confirmed disease event within a specified future interval, linked to a farm/crop observation. Candidate baselines include regularized logistic regression and gradient-boosted trees; temporal neural networks should be considered only when the number and continuity of time series justify them. Weather correlation alone must not be presented as a validated outbreak prediction or causal pathogen model. Temporal evaluation must separate future periods from training periods and group observations by farm where possible.

G. Vegetation and Irrigation Features

Where suitable multispectral imagery exists, vegetation condition may be monitored using the normalized difference vegetation index:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}. \tag{19}$$

NDVI is an indicator of vegetation greenness and stress; it cannot independently identify a specific disease. For irrigation support, the FAO crop-water relationship is

$$ET_c = K_c ET_0, \tag{20}$$

where ET_c is crop evapotranspiration, K_c is the crop coefficient, and ET_0 is reference evapotranspiration [17]. A simplified gross irrigation requirement is

$$I_g = \frac{\max(0,\, ET_c - P_{\mathrm{eff}} - \Delta S)}{\eta_a}, \tag{21}$$

where P_{eff} is effective rainfall, ΔS is the usable root-zone soil-water contribution over the period, and η_a is application efficiency. This estimate requires crop-, soil-, rooting-depth-, and irrigation-specific adaptation before operational use.

H. Constrained Agricultural Recommendation

Let $A_{\mathrm{safe}}(x)$ be the set of actions approved for the crop, growth stage, region, and current evidence by qualified agronomic review. The recommendation layer may rank only actions in this set:

$$a^* = \arg\max_{a \in A_{\mathrm{safe}}(x)} \left[V\, \widehat{\Delta Y}(a \mid x) - C(a) - \lambda E(a) \right], \tag{22}$$

where V is crop value per yield unit, $\widehat{\Delta Y}$ is an estimated yield benefit, $C(a)$ is action cost, and $E(a)$ is an environmental-impact penalty. Unless estimates are supported by suitable field evidence, this formulation is a transparent ranking objective,

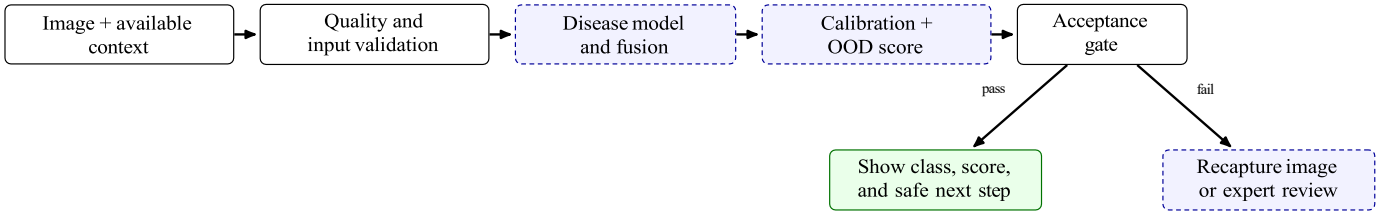


Fig. 2. Selective prediction workflow. Disease output is accepted only when input quality, calibrated confidence, and OOD checks satisfy validation-selected thresholds.

not a causally validated treatment optimizer. Pesticide selection and dosage must never be generated from an image classifier alone; they require verified local guidance and product-label constraints.

V. System Architecture and Implementation

A. Verified Prototype Architecture

The inspected project snapshot implements a Next.js App Router application using React, TypeScript, Tailwind CSS, Next.js route handlers as REST API endpoints, PostgreSQL, and Drizzle ORM. The application includes farmer and administrator workflows for registration and authentication, farm records, crop scans and scan history, follow-up notes, crop-health passport views, alerts, notifications, expert-support requests, reports/CSV export, and administrative oversight. The schema contains user, farm, scan, note, alert, notification, expert-request, and audit-log entities. The user interface provides English, Hindi, and Marathi language options.

The current image-analysis service is a prototype colour heuristic. It resamples an uploaded image to a small canvas, counts broad colour categories, and maps colour ratios to labels such as Healthy, Yellowing, Leaf Blight/Spots, Powdery Mildew, or Uncertain. Its displayed confidence and severity values are heuristic outputs; they are not calibrated probabilities, trained-model scores, or measured accuracy. The current weather function returns region-seeded simulated values and a prototype risk score. The administration endpoint explicitly reports that validated model evaluation data are not available. The prototype README further records that a trained model endpoint, live weather credentials, and production expert network still require integration.

B. Proposed Edge–Cloud Execution

The target architecture separates local and cloud responsibilities. The mobile client should perform image validation and, once a trained compact model is available, lightweight image-only inference using a versioned local artifact. The cloud path may perform contextual fusion, historical analysis, model registry management, and report generation. During disconnection, the client should queue timestamped records and predictions with model version and inference mode; synchronization should be idempotent and preserve whether a prediction was generated locally or remotely. Secure transport, access controls, encrypted storage, and minimum collection of personally identifiable farmer data are deployment requirements, not claims that all such controls are already present.

Algorithm 1 Reliability-Aware Crop-Health Inference

Require: Image I, contextual inputs M, W, S, H, thresholds τ_q, τ_c, τ_d
Ensure: Prediction or defer action
$q \leftarrow \text{ImageQuality}(I)$
if $q < \tau_q$ **then**
  return RequestRecapture
end if
$\mathbf{z}_I \leftarrow f_\vartheta(I)$
for all available contextual modality $j \in \{M, W, S, H\}$ **do**
  compute $\mathbf{z}_j$ and projected embedding $\tilde{\mathbf{z}}_j$
end for
create availability mask $\mathbf{m}$ with $m_I = 1$
compute weights α_j using (8)
$\mathbf{z}_F \leftarrow \sum_j \alpha_j \tilde{\mathbf{z}}_j$
compute logits and calibrated probabilities $\mathbf{p}_T$
compute normalized OOD score d
if $\max_k p_{T,k} < \tau_c$ **or** $d > \tau_d$ **then**
  return ExpertReview($\mathbf{p}_T$, q, d)
end if
$\hat{y} \leftarrow \arg\max_k p_{T,k}$
$A_{safe} \leftarrow \text{ExpertApprovedActions}(\hat{y}, M, W, S)$
return $(\hat{y}, \mathbf{p}_T, A_{safe})$

C. Implementation Boundary and Safety

The existing application can serve as the farmer-workflow and data-management shell, but its heuristic scan output must remain visibly marked as a non-diagnostic prototype until replaced and validated. The model service interface should return at least a class distribution, model identifier, quality/OOD indicators, processing location, and warnings. Expert feedback should be stored with annotator role and timestamp; it should not be automatically treated as ground truth without a validation policy. Any public release should remove demo credentials, use production secrets, and undergo a security review.

VI. Algorithms

Algorithm 1 summarizes the proposed decision flow. The exact quality estimator, OOD detector, calibration threshold, and recommendation rules must be selected and frozen using development data and expert review.

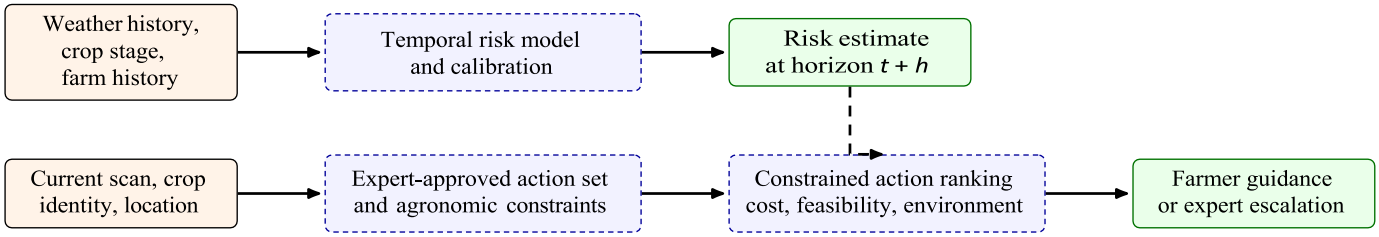


Fig. 3. Proposed risk and recommendation pathway. The forecast informs the decision context, while the recommendation layer ranks only expert-approved actions and preserves an escalation path.

TABLE III
IMPLEMENTATION AUDIT OF THE INSPECTED APPLICATION SNAPSHOT.

Component	Observed status	Evidence / research action
Web application	Implemented prototype	Next.js/React/TypeScript pages and farmer/admin workflows.
API and persistence	Implemented prototype	Next.js route handlers; PostgreSQL via Drizzle ORM; schema for farms, scans, notes, alerts, requests, and logs.
Authentication	Implemented prototype	Password hashing and signed session-token flow; review security configuration before production.
Crop scanning	Prototype heuristic	96-by-96 colour-ratio analysis; not a trained disease classifier. Replace behind existing analysis-service boundary.
Weather and risk	Simulated prototype	Region-derived deterministic demo forecast and unvalidated weighted score; connect timestamped provider data.
Recommendations	Rule-based prototype	Existing advice strings; require crop-, disease-, location-, and label-specific expert review.
Expert workflow	Workflow implemented	Requests and administrative responses; no verified external expert network established by the inspected snapshot.
Reports/admin	Implemented prototype	Summary and CSV export; model-metrics endpoint reports no validated evaluation data.
Offline edge AI	Not verified	Requires a trained, packaged mobile model and on-device latency/memory tests.
Multimodal fusion / calibration	Proposed	Requires feature encoders, training data, calibration split, and test evaluation.
Satellite / IoT integration	Not verified	Requires data provider/device integration and synchronized measurements.

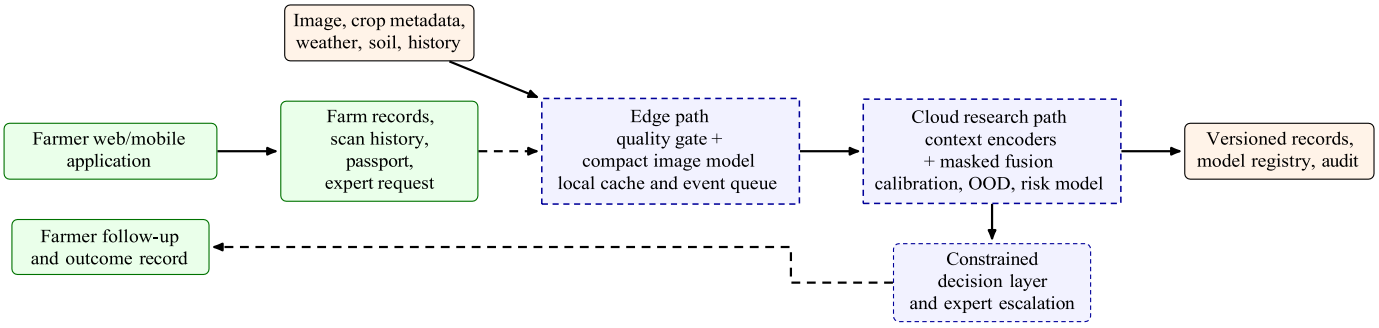


Fig. 4. Proposed edge–cloud execution path. Solid green blocks identify verified application functions; dashed blue blocks are research extensions requiring implementation. The local edge path and later synchronization are target behaviours, not validated deployment results.

VII. EXPERIMENTAL DESIGN

A. Datasets and Label Alignment

Candidate datasets include PlantVillage for controlled-condition image classification, PlantDoc for non-laboratory images, and FieldPlant for plantation-collected images with expert-supervised leaf annotations [3], [6], [7]. Dataset size, crop coverage, class ontology, source licence, annotation procedure, and acquisition conditions must be verified from the version actually used. A locally collected Maharashtra dataset may be added only after obtaining permissions and expert annotation. Public datasets must not be merged merely because their class names look similar: crop species, pathogen labels, disease stages, and healthy/unknown categories require a documented harmonization map.

For every sample, preserve parent-image, leaf, plant, farm, location, and capture-session identifiers when available. Remove exact and near duplicates before splitting. Use grouped training and validation partitions to prevent images of the same plant or capture session from crossing partitions. Reserve a calibration subset separate from model training and a final external field test set untouched by model selection, calibration, or threshold tuning. Cross-domain testing is valid only for classes that can be mapped consistently between source and target datasets. If a class cannot be aligned, report it as an open-set or excluded category rather than silently remapping it.

TABLE IV
CANDIDATE PUBLIC DATASETS AND THEIR INTENDED ROLE. COUNTS MUST BE CHECKED AGAINST THE EXACT DOWNLOADED VERSIONS.

Dataset	Reported scale	Role and caveat
PlantVillage [3]	54,306 images; 14 crops; 38 crop–disease classes	Controlled-image source benchmark; not a field-domain substitute.
PlantDoc [6]	2,598 images; 13 plant species; up to 17 disease classes	Non-laboratory images; audit annotation and class ontology before testing.
FieldPlant [7]	5,170 images; 8,629 leaf annotations; 27 classes	Plantation-collected field data with expert-supervised leaf annotation; useful external benchmark.
Local dataset	To be collected	Requires permission, timestamp/location metadata, expert labels, and documented licensing.

B. Baselines and Ablation Study

The image-only baselines should include a conventional CNN (e.g., ResNet-50), a compact model (e.g., MobileNetV3), an efficient CNN (e.g., EfficientNet-B0), and a transformer baseline where compute permits. The multimodal comparisons should include early concatenation, late fusion, and the proposed masked attention fusion. At minimum, evaluate:

1) image-only classification;
2) image plus crop metadata;
3) image plus weather and soil context;
4) full available-modality fusion;
5) masked fusion under each contextual modality removed in turn;
6) random contextual dropout during training versus no modality dropout;
7) fusion with and without the optional consistency term;
8) calibrated versus uncalibrated prediction;
9) selective prediction at fixed target-risk and fixed-coverage levels;
10) compact edge model versus cloud model under measured device/network conditions.

The missing-modality protocol should test both naturally missing inputs and synthetically masked inputs. Report the mask distribution and ensure synthetic masking does not leak test labels. Compare performance under increasing missingness rates and complete absence of individual contextual channels.

C. Metrics and Statistical Analysis

For classification, report accuracy, macro precision, macro recall, macro F1-score, balanced accuracy, per-class recall, and confusion matrices. For probabilistic outputs, report ECE with its binning scheme, multiclass Brier score, and negative log-likelihood. For selective prediction, report coverage–risk curves, risk at fixed coverage, and coverage at a predefined risk target. For OOD detection, report AUROC and FPR at a specified true- positive rate, with the OOD dataset and score definition stated.

For edge deployment, measure the model artifact size, peak RAM, median and 95th-percentile latency, energy where instrumentation permits, offline inference success, synchronization delay, and behavior during interrupted requests. Name the device, operating system, runtime, model precision, and network conditions. Forecasting evaluation should use temporal separation and report AUROC, AUPRC, Brier score, and lead time. Agricultural outcomes such as yield, pesticide expenditure, water use, or net benefit require a properly designed field study and cannot be inferred from image-classification scores.

Report confidence intervals for key metrics. Use paired comparisons where models are evaluated on the same observations; use a cluster bootstrap at the plant or farm level when observations are grouped. McNemar's test may be used for paired classification errors where its assumptions hold. Apply a multiple-comparison correction such as Holm's procedure when testing several planned hypotheses. Model selection, probability calibration, and threshold selection must not use the final test set.

VIII. RESULTS AND DISCUSSION

The inspected project snapshot does not include a trained disease model, validated weather observations, or an experimental metrics file. Therefore, no model-performance or field-impact result is reported in this draft. This section must be replaced with measured results after implementing and running the experiments.

A. Classification and Multimodal Performance

[EXPERIMENTAL DATA REQUIRED: Insert baseline and multimodal metrics, per-class results, confidence intervals, class-wise confusion matrices, and the grouped field-test protocol.]

B. Calibration, OOD Detection, and Selective Prediction

[EXPERIMENTAL DATA REQUIRED: Insert ECE, Brier score, NLL, OOD metrics, coverage–risk curves, threshold-selection procedure, and accepted/deferred case counts.]

C. Missing-Modality Robustness

[EXPERIMENTAL DATA REQUIRED: Insert results for naturally missing and synthetically masked weather, soil, metadata, and historical inputs. Report the mask distribution and uncertainty intervals.]

D. Edge–Cloud Performance

[EXPERIMENTAL DATA REQUIRED: Insert device model, operating system, runtime, model size, median/P95 latency, RAM, energy if measured, network conditions, and synchronization behaviour.]

E. Discussion

Interpret results against RQ1–RQ5. Distinguish a gain in classification accuracy from a gain in calibration or selective risk. A model that abstains frequently may reduce accepted-prediction error while providing little coverage; both quantities must be discussed. Similarly, a cloud model may improve accuracy but be unsuitable for low-connectivity settings if its availability or latency is poor. No causal claim about yield

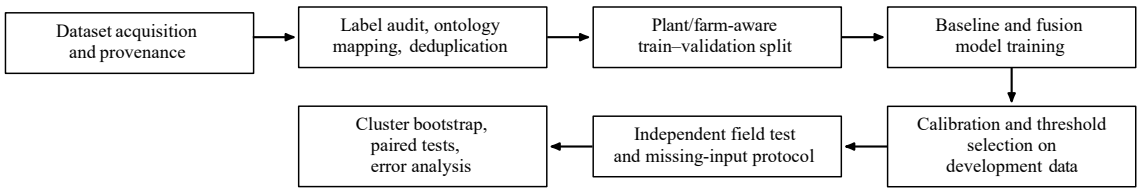


Fig. 5. Leakage-resistant experimental workflow. The final field test is isolated from model selection, calibration, and operating-threshold tuning.

improvement or input reduction should be made without a controlled field study.

IX. Limitations and Threats to Validity

The proposed framework depends on the quality, coverage, and temporal alignment of crop labels and contextual measurements. Public datasets may not represent regional cultivars, disease prevalence, camera devices, or field conditions in Maharashtra. Differences in class ontologies may prevent direct cross-dataset comparison. Leakage can arise from near-duplicate images or repeated images from the same plant, farm, or session. Calibration can deteriorate under unseen domain shift, and an OOD score is not a guarantee of safe recognition. Weather associations do not by themselves establish disease causation. NDVI measures vegetation condition rather than pathogen identity. Incorrect agricultural recommendations can cause economic or environmental harm; therefore, high-impact interventions require expert-reviewed rules, local product-label constraints, and an escalation path. Finally, the current application analysis and weather services are demonstration implementations, not validated AI or sensing subsystems.

X. Conclusion and Future Work

This paper has specified *Dharti Mitra* as a reliability-centred multimodal edge–cloud framework for crop-health monitoring and agricultural decision support. The methodology defines availability-aware feature fusion, independent probability calibration, explicit selective acceptance conditions, temporal risk forecasting, and constrained recommendation ranking. The source-code audit shows that the current application provides farmer workflows and data management, while image analysis remains a colour heuristic and weather data are simulated. The proposed neural and edge components therefore require implementation and empirical validation. Future work should train and evaluate image baselines, add expert-labelled contextual observations, test missing- modality handling and field-domain generalization, and benchmark on named mobile hardware under realistic connectivity conditions. A prospective field study is required before claiming changes in crop yield, pesticide use, water consumption, or farmer income.

Acknowledgment

[AUTHOR ACTION: Insert confirmed project-guide, field-partner, dataset, and funding acknowledgments, or remove this section if none apply.]

References

[1] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming—a review," *Agricultural Systems*, vol. 153, pp. 69–80, 2017, doi: 10.1016/j.agsy.2017.01.023.

[2] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, Art. no. 1419, 2016, doi: 10.3389/fpls.2016.01419.

[4] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.

[5] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019, doi: 10.1016/j.compag.2018.03.032.

[6] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, "PlantDoc: A dataset for visual plant disease detection," in *Proc. 7th ACM IKDD CoDS and 25th COMAD*, 2020, pp. 249–253, doi: 10.1145/3371158.3371196.

[7] E. Moupojou *et al.*, "FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning," *IEEE Access*, vol. 11, pp. 35398–35410, 2023, doi: 10.1109/ACCESS.2023.3263042.

[8] K. Xiang, D. Shi, and X. Zhu, "Quantifying the reliability gap in cross-domain plant disease classification: Benchmarking the limited efficacy of standard mitigation techniques under controlled-to-field shift," *Frontiers in Plant Science*, vol. 17, Art. no. 1826962, 2026, doi: 10.3389/fpls.2026.1826962.

[9] T. Shrivastava, A. K. Agrawal, A. Shukla, and P. Dubey, "An uncertainty-aware evaluation framework based on hierarchical vision transformers for robust cross-domain plant leaf disease classification," *Scientific Reports*, vol. 16, Art. no. 24348, 2026, doi: 10.1038/s41598-026-55107-6.

[10] N. Zakeeya and S. L. Sugali, "Context aware multimodal plant disease diagnosis using agro ecological metadata and explainable inference," *Discover Computing*, vol. 29, Art. no. 734, 2026, doi: 10.1007/s10791-026-10536-9.

[11] X. Wang and J. Liu, "Enhancing plant disease detection through multi-modal integration of visual and textual data," *Plant Methods*, vol. 22, Art. no. 30, 2026, doi: 10.1186/s13007-026-01521-w.

[12] M. G. F. Odounfa and R. L. Glèlè Kakaí, "Multimodal multitask deep learning for chili pepper disease diagnosis and severity estimation using images and agrometeorological data," *Scientific Reports*, vol. 16, Art. no. 37619, 2026, doi: 10.1038/s41598-026-64732-0.

[13] T. Manda, T. Huang, Y. Ding, S. Dai, L. Yang, and T. Dai, "Multimodal deep learning and foundation models for early detection and forecasting of plant diseases," *Plants*, vol. 15, no. 17, Art. no. 2564, 2026, doi: 10.3390/plants15172564.

[14] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th International Conference on Machine Learning*, PMLR, vol. 70, 2017, pp. 1321–1330.

[15] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[16] Y. Ovadia *et al.*, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[17] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. Rome, Italy: Food and Agriculture Organization of the United Nations, 1998, FAO Irrigation and Drainage Paper 56.