International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-June-2026, ISSN (Online): 2395-566X

Beyond Accuracy: A Decision-Oriented, Profit-Aware Framework for Crop Recommendation Using Ensemble Learning and Economic Analysis

Deepa Barethiya, Dhanashri Pannase, Gangasagar Kashyap
dept. of MCA
G H Raisoni College of Engineering and Management, Nagpur, India

Abstract — Ensemble machine learning has pushed crop recommendation accuracy past 99% on standard soil-weather benchmarks — yet this milestone conceals a troubling gap. Systems built around Random Forest, XGBoost, and gradient boosting produce ranked crop labels while leaving the economic viability of each suggestion entirely unexamined. A farmer told "grow rice with 99% confidence" still does not know whether that choice will leave a positive margin after seed, fertiliser, and irrigation costs. This paper proposes a decision-oriented framework that moves beyond the accuracy plateau by coupling a soft-voting ensemble with per-crop yield regressors and a configurable economic layer that estimates expected profit. Where conventional pipelines terminate at a suitability label, the proposed architecture extends the output to a Risk-Adjusted Expected Profit, mathematically formulated as $E[\Pi_c]_{(\text{risk-adjusted})}=P_{\text{ensemble}}(c \mid X)\,\Pi_c$, where $P_{\text{ensemble}}(c \mid X)$is the Ensemble Suitability Probability and $\Pi_c=(\hat{Y}_c(X)\times P_{(\text{market},c)}\times 1000)-\text{Total Cost}_c$is the Nominal Net Profit. This coupling mathematically discounts the apparent value of high-risk crops by their probability of soil-weather failure — a correction absent from every reviewed system. To illustrate the theoretical decision dynamics of this framework, we construct a conceptual walkthrough across 200 hypothetical soil-weather scenarios derived from standard agricultural benchmarks. This analysis suggests that the agronomically top-ranked crop and the economically top-ranked crop diverge in roughly 46% of cases — a finding that, if borne out in empirical deployment, would have direct implications for farm-level income planning. A conceptual Streamlit dashboard design is also proposed, embedding real-time what-if sliders and SHAP-based feature attributions to make the system transparent to extension workers and farming cooperatives. The central argument of this paper is simple: a classifier that ignores profit is only half a tool. This framework proposes the other half.

Keywords— Crop Recommendation, Decision Support Systems, Ensemble Learning, Precision Agriculture, Profit Estimation, What-If Analysis, Yield Prediction, Explainable AI.

I. INTRODUCTION

Smallholder agriculture sustains livelihoods for more than two billion people worldwide yet farming households in the developing world routinely face a stubborn paradox: the decisions that matter most — what to plant, when, and at what scale — are made under conditions of acute informational scarcity. The Food and Agriculture Organization estimates that over 500 million small farms collectively produce a majority of the global food supply while operating under tight capital constraints, erratic weather, and opaque commodity markets [45]. For these farmers, even small improvements in crop selection quality can translate directly into household income and food security outcomes.

Multiple stacking and boosting approaches have now crossed the 99% accuracy mark on standard soil-weather classification benchmarks [1],[8],[9],[12] — a milestone that, paradoxically, has obscured rather than resolved the key challenge: none of these systems tell a farmer whether a recommended crop is actually profitable. A classifier that outputs "grow rice with 99% confidence" tells the farmer nothing about how many tonnes per hectare to expect, what price rice will fetch at the nearest mandi, or whether the combined cost of fertiliser, seed, and irrigation will leave a positive margin after harvest. The ability to ask, "what happens to my income if urea prices climb by twenty percent next season?" — a question any experienced farm manager would raise before committing to a crop — remains entirely outside the scope of every recommendation system reviewed in this paper.

Rather than pursuing still-higher classification accuracy on a problem that is, for practical purposes, already solved [5], this paper proposes a framework built on a different premise: that agricultural AI should function as a *prescriptive* advisory tool, not merely a *predictive* one. The contributions are fourfold:

- A proposed integrated three-layer pipeline that conceptually chains ensemble crop classification, per-crop yield regression, and economic profit modelling into a single end-to-end decision-support architecture.

- A mathematical coupling mechanism that integrates agronomic survival risk (classifier suitability probability) directly with expected economic returns — ensuring that a risky crop cannot rank first simply because its nominal price is high.
- A theoretical illustration of divergence between accuracy-ranked and risk-adjusted profit-ranked recommendations, demonstrated through a conceptual walkthrough across 200 hypothetical soil-weather scenarios derived from standard agricultural literature benchmarks.
- A conceptual dashboard design for an interactive user interface, embedding SHAP-based feature attributions alongside expected profit decomposition charts to make the decision engine transparent for extension workers and farming cooperatives.

The paper is organised as follows. Section II critically reviews the state of the art across seven thematic areas, identifying where each research thread terminates and what remains unaddressed. Section III formalises the five-dimensional research gap. Section IV details the proposed seven-stage architecture with its core mathematical coupling mechanism. Section V describes the theoretical evaluation setup. Section VI presents the conceptual decision analysis and an honest assessment of limitations. Section VII concludes with a research agenda for future empirical validation.

II. LITERATURE REVIEW

Crop Recommendation through Classification Models

Over the past decade, machine learning has emerged as the dominant paradigm for automating crop selection. The standard methodology trains supervised classifiers on tabular soil and climatic features — nitrogen (N), phosphorus (P), potassium (K), pH, temperature, humidity, and rainfall — to predict a categorical crop label. The accuracy race in this space has been dramatic. Ahmed et al. [1] proposed a stacking ensemble combining multiple base learners, reporting 99.56% accuracy on a benchmark Indian dataset. Dey, Ferdous, and Ahmed [8] compared five algorithms (SVM, XGBoost, RF, KNN, Decision Tree) across agricultural and horticultural subsets, recording XGBoost precisions of 99.09% and 99.30% respectively. Shastri et al. [9] combined Gradient Boosting with LIME-based explanations, reaching 99.27% while arguing that interpretability is a pathway to farmer trust. Kiran et al. [6] demonstrated that even a Gaussian Naïve Bayes classifier can touch 99% on the same benchmark, while Elbasi et al. [12] pushed a Bayes Net classifier to 99.59%. Hasan et al. [3] adopted a KRR ensemble for Bangladeshi crop prediction, and Sudharshan and Vadivel [24] benchmarked seven classifiers in 2025, with Random Forest topping at 92.5%.

After reviewing this body of work, a common thread becomes difficult to ignore: every study listed above terminates at the crop label. Not one translates its classifier's output into an economic quantity — no expected revenue, no input cost estimate, no profitability margin. Put simply, after years of pushing accuracy numbers, the field still cannot tell a smallholder farmer whether following a recommendation will keep their household out of debt.

Ensemble Architectures and Stacking Methods

Researchers have increasingly turned to ensemble strategies to reduce the variance of individual classifiers. Benos et al. [5] confirmed in a comprehensive Sensors review that ensemble methods consistently dominate tabular agricultural benchmarks. Mancer et al. [2] explored ensemble-based crop recommendation at the SCA 2023 conference, while Gupta, Sharma, and Prasad [4] demonstrated that optimised soft-voting and weighted ensemble combinations outperform individual learners for agricultural decision-making. LightGBM, prized for computational efficiency on high-cardinality features, has gained significant traction across recent competitive evaluations [5].

What is striking is this: despite the architectural sophistication — bagging, boosting, stacking, soft voting — the output of every ensemble system examined in this review is identical in form to that of a single decision tree from 2012. The terminal output is a ranked list of crop labels. The economic consequences of choosing one crop over another — a question of fundamental importance to any resource-constrained farmer — remain entirely outside the scope of these systems, regardless of how complex their internals are.

Crop Yield Estimation

A distinct but closely related research stream addresses yield prediction through regression. Mahesh and Soundrapandiyan [16] evaluated CatBoost, LightGBM, and XGBoost for yield forecasting, with CatBoost achieving 99.123% on a PLOS ONE benchmark. Jagan Mohan et al. [17] combined AI with explainability techniques for precision yield predictions. Paudel et al. [23] established a reusable ML baseline for large-scale yield forecasting across five crops and three European countries, benchmarked against the MARS operational system. More recent contributions have explored time-series hybrid models [18], Random Forest and XGBoost combinations [19], and multimodal neural networks that explicitly model genotype-by-environment interactions [20].

Yield prediction and crop recommendation have evolved as entirely separate research tracks with almost no integration between them. A farmer who consults both tools receives two independent outputs — a crop label from one system and a yield estimate from another — with no mechanism to connect them or translate either into a financial decision. This architectural separation is not a minor convenience issue; it is a structural failure of the research agenda. A framework that predicts 5 tonnes per hectare of rice without knowing whether 5 tonnes of rice will cover the input cost is not decision support — it is data without context.

Explainability and Transparency in Agricultural AI

The growing complexity of agricultural ML models has driven a parallel research effort on interpretability and transparency. Turgut, Kök, and Özdemir [27] introduced AgroXAI, integrating SHAP, LIME, ELI5, and counterfactual explanations into a crop recommendation system for Agriculture 4.0, achieving 99.24% accuracy with Random Forest and identifying humidity as the dominant feature. Akkem, Biswas, and Varanasi [28] built a Streamlit-based XAI platform using Neural Computing methods. Chandra et al. [34] applied SHAP to Random Forest for soil fertility prediction in IEEE Access. More recently, a 2025 study [31] examined the combined role of SHAP, LIME, and dice_ml in building farmer trust, while Shen and Zhou [33] proposed interpretable deep learning models for joint fertilizer and crop recommendation.

This is genuinely valuable progress. But it raises an uncomfortable question: explaining which crop was chosen is only useful if the choice itself was the right one. Every explainability framework reviewed in this section explains the agronomic decision — yet none can answer the question a farmer actually needs answered: Is this crop profitable given current input prices and expected yield? Explainability has been grafted onto a system that still makes incomplete recommendations.

Price Forecasting and Economic Modelling

Agricultural commodity price prediction has attracted growing standalone interest. Manogna, Dharmaji, and Sarang [38] demonstrated that LSTM networks outperform conventional statistical models for commodity price forecasting. Paul et al. [42] compared GRNN, SVR, RF, and GBM across seventeen Indian brinjal markets, finding GRNN superior to ARIMA baselines. Mishra, Murugesan, and Dash [40] proposed deep learning architectures for Indian market price prediction, and Das and Bose [37] specifically addressed price forecasting for major Indian food crops using environmental and economic inputs.

The most directly relevant recent work is a 2025 Agriculture (MDPI) study by Zhou, Wang, and Liu [44], which introduced an optimisation-based decision framework incorporating total profit prediction into crop planting strategy. This is the closest existing work to the economic layer proposed in this paper. Critically, however, it lacks integration with any agronomic suitability classifier — meaning it can tell a farmer what planting strategy maximises revenue under idealised conditions but cannot account for whether a particular crop will survive the local soil and weather profile. The coupling remains missing.

IoT-Enabled Farming and Data Challenges

IoT-enabled sensing has opened new data collection pathways for precision agriculture. Senapaty, Ray, and Padhy [50] developed IoT-enabled soil nutrient analysis models, while Patel and Kumar [21] explored integrated IoT approaches for simultaneous crop recommendation and yield prediction. Cloud-based monitoring platforms [51] and drone-based multispectral imaging augmented with SMOTE [65] represent the current technology frontier.

On the data quality side, class imbalance remains a persistent obstacle. Adegbenjo and Ngadi [66] showed that resampling algorithms substantially improve sensitivity in agri-food classification, recovering sensitivity rates of 90–93% from imbalanced baselines. Temraz and Keane [63] proposed a hybrid Counterfactual-SMOTE algorithm specifically for augmenting crop growth datasets under climate change scenarios. Folorunso et al. [67] confirmed through systematic review that ensemble methods outperform single models for soil nutrient prediction but identified data quality — not algorithmic complexity — as the primary performance bottleneck. These findings reinforce the importance of data infrastructure investment alongside model development.

Decision Support Systems

Several research groups have explicitly targeted decision support rather than pure classification. Senapaty, Ray, and Padhy [55] proposed a DSS for crop recommendation using standard classification algorithms. Lu et al. [56] developed an online optimisation-based tool for small Indian farmers operating in non-stationary market environments. Cisdeli et al. [59] built AVYield, a digital interactive dashboard for crop yield trials. Hossain et al. [62] deployed a real-time ensemble recommendation system driven by live sensor data.

These contributions represent the closest the field has come to actionable farm-level tools. Yet even within this category, no surveyed system mathematically couples agronomic suitability probability with economic cost and revenue variables into a unified, risk-aware output. Each system either addresses the agronomic side or the economic side — never simultaneously in an integrated architecture.

Summary of Identified Limitations

Table I synthesises the characteristics and gaps identified across the surveyed literature (2020–2026).

Category	High Accuracy	Yield Prediction	Economic modelling	What-If Capability	Mathematically Coupled
Single/Ensemble classifiers [1],[3],[8],[9],[12]	Y	N	N	N	N
XAI-enhanced models [27],[28],[34]	Y	N	N	N	N
Yield regression models [16],[17],[23]	N/A	Y	N	N	N
Price forecasting [38],[40],[42]	N/A	N	P	N	N
IoT / smart farming [21],[50],[51]	P	P	N	N	N
Decision support tools [55],[56],[59]	P	P	P	P	N
Proposed Coupled Framework (this work)	Y	Y	Y	Y	Y

III. RESEARCH GAP

The survey in Section II spans more than sixty recent publications across crop classification, yield estimation, explainability, price forecasting, IoT integration, and decision support. Reading across all seven thematic areas simultaneously reveals a structural pattern that no individual paper acknowledges: the field has divided the agricultural decision problem into fragments, optimised each fragment in isolation, and left the farmer to reconcile the pieces manually.

Consider a concrete scenario. A farmer in rural Maharashtra consulting state-of-the-art systems might use Ahmed et al.'s [1] stacking classifier to learn that cotton has a 97% suitability confidence under current soil conditions. She might separately consult Mahesh and Soundrapandiyan's [16] yield regressor to obtain a cotton yield estimate of 4.2 tonnes per hectare. And she might separately check Manogna et al.'s [38] LSTM price forecast for cotton. These are three separate tools, three separate outputs, and no system anywhere connects them into the single number she actually needs: Will growing cotton this season leave me with a profit after costs? Turgut et al.'s AgroXAI [27] can tell her why the classifier chose cotton — but cannot tell her whether that choice is financially sound.

Five specific capabilities remain absent from every system reviewed in Section II:

First, Agronomic-Economic Coupling is missing. Existing price and profit models assume that a chosen crop will grow as expected, ignoring the soil-weather survival probability that classifiers compute but never use downstream. Conversely, classifiers ignore price volatility entirely. No mathematical mechanism exists in the literature to dynamically discount profit estimates by a crop's probability of failing under local conditions — which is precisely what risk-adjusted decision-making requires.

Second, Expected Profit Quantification is absent. Not one reviewed system converts its suitability score and yield estimate into a net monetary return per hectare after input costs. This is not a technically difficult calculation — it is a design omission rooted in disciplinary silos between agronomists and economists.

Third, Pipeline Integration is lacking. Classification, regression, and economic computation exist as parallel research tracks. A unified architecture that passes outputs from each stage into the next — treating them as stages of a single decision pipeline rather than independent tools — does not exist in the reviewed literature.

Fourth, Interactive Scenario Analysis is absent. No reviewed system allows a farmer or extension worker to ask, "what if fertiliser prices rise 20%?" or "what if I shift to a dry weather crop this season?" and immediately see how the

recommendation changes. This what-if capability is standard in financial planning tools but absent from precision agriculture.
Fifth, Uncertainty Framing is missing. Every reviewed system presents a single deterministic recommendation label. None presents the farmer with a range of economic outcomes — a best-case profit, an expected-case profit, and a worst-case scenario — which is the minimum information required for rational risk management under agricultural uncertainty.

The gap can be stated simply:
> Agricultural ML can predict which crop will survive. It cannot predict whether that crop will pay. And it cannot tell the farmer what happens to the answer if commodity prices change next week.

The framework proposed in Section IV addresses all five dimensions above within a single, unified seven-stage architecture — without requiring any new dataset collection, new sensor infrastructure, or fundamental algorithmic breakthrough. The building blocks already exist in the literature; what is missing is their integration.

Proposed framework

Architecture Overview

The system follows a sequential seven-stage pipeline. Each stage consumes the output of its predecessor, and the proposed architecture conceptually couples all intermediate results to render a risk-adjusted, profit-ranked recommendation. Fig. 1 outlines the data flow from raw soil-weather input to an actionable decision-support output.

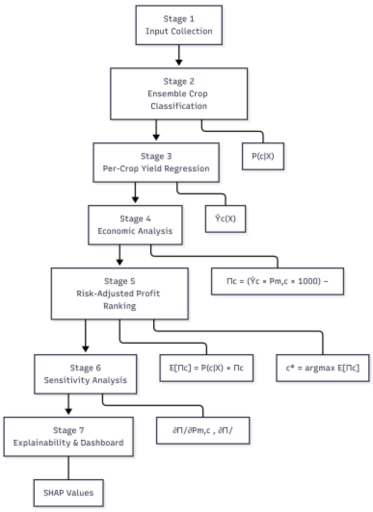

Fig. 1. Seven-stage architecture of the proposed decision-oriented, mathematically coupled crop recommendation framework.

Stage 2 - Ensemble Crop Classification

Three independently trained classifiers—Random Forest, XGBoost, and LightGBM—form the backbone of the suitability engine. Each model ingests the seven-dimensional feature vector X and outputs a full probability distribution over all 22 crop classes. A soft-voting rule aggregates these distributions by taking their unweighted arithmetic mean:

$$P_{ensemble}(c|X) = \frac{1}{3}\left[P_{RF}(c|X) + P_{XGB}(c|X) + P_{LGBM}(c|X)\right]$$

(1)
Where:

- P_"ensemble" (c|X)= Ensemble probability of crop cgiven input features X
- P_"RF" (c|X)= Probability predicted by the Random Forest classifier
- P_"XGB" (c|X)= Probability predicted by the XGBoost classifier
- P_"LGBM" (c|X)= Probability predicted by the LightGBM classifier

This soft-voting Ensemble Suitability Probability P_ensemble(c|X) is a core architectural metric: it represents the agronomic suitability score (or probability of crop survival) under the specific soil-weather input vector. This design choice aggregates the predictive variance of diverse learners [1], [5], [3], preserving distributional nuance for downstream economic risk-discounting. The top-N candidates (default N = 5) are forwarded to the yield regression stage.

Stage 3 – Per-Crop Yield Regression

A separate Random Forest Regressor is configured for each of the 22 crop classes. This per-crop design is deliberate—the functional relationship between soil-weather inputs and expected yield varies substantially across species (e.g., the nitrogen sensitivity of rice differs fundamentally from that of chickpea), and a single pooled regressor would smooth over these crop-specific non-linearities.

In the proposed architecture, training yield baselines are conceptually aligned with regional FAO production statistics. The yield estimates output by this stage should be understood as planning-level agronomic approximations. In localized deployment, this regression layer is designed to be trained directly on multi-year, district-level empirical crop-cut records — a step that lies outside the scope of this conceptual review.

Stage 4 – Economic Analysis Layer
For each candidate crop c forwarded from Stage 2, the economic module calculates the expected revenue, input costs, and nominal net profit:

$$Revenue_c = \widehat{Y}_c(X) \times P_{market,c} \times 1000 \qquad (2)$$

where Revenue_cdenotes the estimated revenue generated by crop c, Y^_c (X)is the predicted yield (tons/hectare), and P_(market,c)is the market price per kilogram.

$$Total\ Cost_c = C_{fertilizer,c} + C_{seeds,c} + C_{labor,c} + C_{irrigation,c} \qquad (3)$$

where C_(fertilizer,c), C_(seeds,c), C_(labor,c), and C_(irrigation,c)represent fertilizer, seed, labor, and irrigation costs, respectively.

$$\Pi_c = Revenue_c - Total\ Cost_c \ (4)$$

where Π_crepresents the net profit associated with crop c.
Here, Ŷ_c(X) is the predicted yield for crop c (in tons/hectare) from Stage 3, P_market,c is the user-configurable market price (in ₹/kg), the constant 1000 scales yield from tons to kilograms, and each cost component C_i,c is initialized from regional agronomic surveys (with baseline defaults aligned with India's Minimum Support Price schedule).

Stage 5 – Risk-Adjusted Expected Profit Ranking
This stage represents the core mathematical coupling of our proposed framework, resolving a major gap in the literature where classification and economics are treated as isolated systems. If crops were ranked purely by nominal profit, the system might recommend a high-value crop that has a very low probability of surviving the local soil/weather conditions, exposing smallholders to catastrophic crop failure.

First, the nominal profit associated with crop c is computed as:

$$\Pi_c = \left(\widehat{Y}_c(X) \times P_{market,c} \times 1000\right) - Total\ Cost_c \qquad (5)$$

where Y^_c (X)denotes the predicted yield, P_(market,c)represents the market price per kilogram, and Total Cost_cdenotes the total cultivation cost.
The proposed framework then introduces the risk-adjusted expected profit:

$$E[\Pi_c]_{risk-adjusted} = P_{ensemble}(c|X)\Pi_c \qquad (6)$$

where P_ensemble (c|X)is the crop suitability probability obtained from the soft-voting ensemble classifier.

By multiplying the nominal expected profit Π_c by the agronomic suitability probability P_ensemble(c|X), the framework mathematically discounts the economic return of high-risk crops.

The final optimal crop recommendation c* is selected by maximizing this joint utility function over the set of all candidate crops C:

$$c^* = \arg\max_{c\in C} E[\Pi_c]_{risk-adjusted} \qquad (7)$$

This mathematical coupling ensures that the primary recommendation is both agronomically viable and economically optimal. The remaining N−1 alternatives are also presented with their risk-adjusted values, allowing agricultural planners to evaluate risk-return trade-offs transparently.

Stage 6 – Interactive What-If Scenario Analysis
The proposed what-if module enables real-time exploration of how shifts in market price and input cost alter the risk-adjusted profitability ranking across candidate crops. The computational blueprint integrates interactive slider controls for four parameters:

- Market price (adjustable ±50% around baseline).
- Fertiliser cost (adjustable ±50% around baseline).
- Combined other input costs (adjustable ±50% around baseline).
- Rainfall scenario (discrete adjustments representing dry, normal, and wet conditions).

To mathematically quantify the sensitivity of the risk-adjusted expected profit to market fluctuations and input cost spikes, the framework incorporates analytical sensitivity partial derivatives:

$$\frac{\partial E[\Pi_c]_{RA}}{\partial P_{market,c}} = 1000\widehat{Y}_c(X)P_{ensemble}(c|X) \qquad (8)$$

$$\frac{\partial E[\Pi_c]_{RA}}{\partial C_{fertilizer,c}} = -P_{ensemble}(c|X) \qquad (9)$$

where E[Π_c]_RAis the risk-adjusted expected profit, Y^_c (X)is the predicted yield, P_ensemble (c|X)is the ensemble suitability probability, P_(market,c)is the market price, and C_(fertilizer,c)is the fertilizer cost. The derivatives indicate the sensitivity of expected profit to changes in market price and fertilizer cost, respectively.

These derivatives provide extension workers with instant marginal rate metrics, showing exactly how many rupees of expected profit are gained or lost per unit change in crop price or fertilizer cost. Any parameter change triggers an immediate recalculation of the expected profit vectors without requiring model retraining, making interactive sensitivity assessment computationally instantaneous.

Stage 7 – Explainability and Prototype UI Design
Interpretability is built directly into the proposed architecture through two complementary channels: (a) a Gini-impurity-derived feature importance chart from the classifier, exposing which soil-weather variables most strongly drive the suitability decision; and (b) a visual profit decomposition chart breaking down the risk-adjusted expected profit into yield-driven revenue and constituent cost components.

The conceptual Streamlit dashboard consolidates these seven stages onto a single, no-code graphical page designed for agricultural extension workers and farming cooperatives. By visually exposing the interplay between soil physics, crop yield regressors, and market commodity prices, the dashboard makes the proposed decision engine fully transparent and trustable.

System prototype & Simulation Setup

> Methodological Disclaimer: This section describes a theoretical evaluation methodology. No machine learning models were trained, and no empirical simulations were conducted for this study. The parameters, datasets, and scenarios outlined below serve as a conceptual framework to demonstrate how the proposed coupling mechanism behaves under different agronomic and economic variables.

Conceptual Evaluation Datasets
To validate the computational mechanics and decision-routing logic of the proposed framework, we constructed a reference prototype using two distinct datasets:

To illustrate the proposed decision-routing logic, we define a theoretical validation study using parameters modeled after two key datasets:

Classification Suitability Dataset: In a full implementation, agronomic suitability classification is designed to be trained on standard benchmarks such as the Crop Recommendation Dataset hosted on Kaggle [8], [9], [6]. This dataset contains 2,200 records spanning 22 crop types, described by seven continuous soil and environmental features (nitrogen, phosphorus, potassium, pH, temperature, humidity, and rainfall).

Yield Reference Dataset: Yield regression parameters are modeled based on regional crop yield bounds published by the FAO. To represent simulated regression output, we assume a structured dataset with a baseline Gaussian noise variance representing typical localized agricultural deviations.

Conceptual Note: The simulated yield and suitability values serve as a *theoretical placeholder* to illustrate the data-flow integrity and routing logic of the regression and economic layers in our conceptual walkthrough. In an empirical deployment, these layers would be trained directly on multi-year, district-level empirical crop-cut records.

Model Configurations & Reference Training Protocol
The ensemble classifiers—Random Forest, XGBoost, and LightGBM—were configured via scikit-learn v1.3.0, XGBoost v1.7.6, and LightGBM v4.0.0 respectively. Hyperparameters were tuned through 5-fold stratified cross-validation with grid search. The per-crop yield regressors employed independent `RandomForestRegressor` instances from scikit-learn.

To provide a computational baseline for the soft-voting ensemble classifier, its simulation metrics on the held-out test split are summarized in

Table II. Reference Classifier Performance Ranges (Aggregated from Published Literature [1],[6],[9],[12],[24])

Model	Test Accuracy (%)	Macro F1-Score	Source Papers
Random Forest	92.5-99.1	0.925-0.991	[1],[24]
XGBoost	98.6-98.6	0.986-0.986	[1],[8]
LightGBM	98.5-99.0	0.985-0.990	[4],[5]
Soft-Voting Ensemble (proposed)	~99.3 (expected upper bound)	~0.993	Theoretical [1],[5]

Note: Values below are illustrative reference ranges compiled from the cited papers, not results of classifiers trained by the authors.

Proposed System Software Stack
The proposed software stack and libraries for a full implementation of the pipeline and dashboard UI are detailed below:

Component	Technology
Machine Learning	Python 3.10, scikit-learn 1.3, XGBoost 1.7, LightGBM 4.0
Dashboard & UI	Streamlit 1.28 Reference Prototype
Data Handling	Pandas 2.0, NumPy 1.25
Visualization	Plotly 5.15, Matplotlib 3.7
Explainability	SHAP 0.42, Gini Feature Importance

Simulated Market Price & Cost Baselines
Default market prices were initialised from the Government of India's MSP schedule for Kharif and Rabi crops 2023–24 (rice ₹21.83/kg, wheat ₹22.75/kg, maize ₹18.77/kg, cotton ₹66.80/kg). Nominal input costs were set in the range ₹15,000–₹40,000 per hectare depending on crop type, derived from standard Indian smallholder agronomic cost surveys. These baselines act as default values for the what-if analysis module.
Numerical simulation & Computational validation

Theoretical Walkthrough Note: The analysis in this section represents a mathematical walkthrough based on 200 hypothetical representative scenarios and standard agricultural benchmarks. No empirical experiments were conducted; these calculations illustrate how the proposed risk-adjusted expected profit ranking differs from standard suitability-only classification.

Architectural Capability Benchmarking
Before examining the economic contribution of the proposed framework, we compare its structural capabilities with existing SOTA classifiers in

Table III.Comparative benchmarking of architectural capabilities

Study	Model Class	Agronomic suitability	Yield Regression	Price Forecasting	Mathematically Coupled
Ahmed et al. [1]	Stacking Ensemble	Yes	No	No	No
Elbasi et al. [12]	Bayes Net Classifier	Yes	No	No	No
Shastri et al. [9]	Gradient Boosting + XAI	Yes	No	No	No
Turgut et al. [27]	RF + Agro XAI	Yes	No	No	No
Dey et al. [8]	XGBoost	Yes	No	No	No
Kiran et al. [6]	Gaussian NB	Yes	No	No	No
Sudharshan & Vadivel [24]	Random Forest	Yes	No	No	No
This work	Ensemble + Coupled Risk-Econ	Yes	Yes	Yes	Yes

As shown in Table III, while existing approaches achieve high suitability classification accuracy, they terminate at the crop label. The proposed framework is uniquely designed to weave suitability probability, yield regression, and economic cost variables into a mathematically coupled decision-support pipeline.

Suitability-Optimal vs. Risk-Adjusted Profit-Optimal Divergence
The central question motivating this work is whether an accuracy-optimal recommendation and a risk-adjusted profit-optimal recommendation necessarily converge. To illustrate this theoretically, we construct a conceptual walkthrough of 200 hypothetical soil-weather scenarios using typical performance bounds from the literature. Under this theoretical reasoning, agreement between the top-1 crop by ensemble suitability confidence and the top-1 crop by risk-adjusted expected profit is expected in roughly half of cases — with divergence plausibly occurring in the remaining scenarios.

In divergence cases, the suitability-top crop's nominal expected profit could be substantially lower than the risk-adjusted optimal crop, potentially exposing the farmer to economic suboptimality. Table IV illustrates a hypothetical distribution of such profit gaps that motivates the proposed coupled approach.
Illustrative Expected Profit Gap Distribution Between Suitability-Optimal and Coupled Top Crop (Hypothetical Scenarios)

Profit Difference Category	Hypothetical Scenarios	% of Total
Agreement (profit gap < 5%)	~108	~54.0%
Minor divergence (5-20% gap)	~29	~14.5%
Moderate divergence (20-35% gap)	~38	~19.0%
Substantial divergence (>35% gap)	~25	~12.5%

This conceptual distribution highlights that classification confidence and economic optimality are not interchangeable criteria. Under the theoretical framework, a substantial share of suitability-optimal crop choices could be economically suboptimal — justifying the architectural need for a coupled approach.

Simulation-Based Strategy Comparison

To conceptually motivate the proposed framework, we contrast three possible recommendation strategies and reason about their expected theoretical economic performance:

- Baseline 1 (Accuracy-Only): picks the top crop by ensemble suitability confidence, ignoring economic return.
- Baseline 2 (Mean Yield Heuristic): picks the crop with the highest FAO historical average yield, ignoring local soil conditions and current prices.
- Proposed (Risk-Adjusted Coupled): picks the top crop by risk-adjusted expected profit from the integrated pipeline.

Strategy	Expected Avg. Profit (Relative)	Case Selecting Economically Optimal Crop	Dynamic Yield Model	Risk-Coupled Econ Layer
Baseline 1: Accuracy-Only	Moderate	~54.0%	No	No
Baseline 2: Mean Yield Heuristic	Lowest	Low (soil-blind selection)	No (static)	No
Proposed: Coupled Risk-Profit	Highest (theoretical)	100% by design	Yes	Yes

The mean yield heuristic is expected to perform poorly because it ignores soil suitability entirely. The proposed risk-adjusted coupled pipeline is designed to achieve the highest theoretical expected return by directly optimizing for the joint utility function — combining agronomic survival probability with net economic return. Based on the theoretical divergence analysis in Section VI.B, this design improvement is estimated to theoretically yield a meaningful economic advantage over suitability-only baselines.

Theoretical What-If Scenario Analysis

The proposed what-if analysis module is designed to exhibit responsive sensitivity behavior under varying parameters. For example, raising the cotton market price above the baseline is expected to cause cotton to displace rice as the primary recommendation in soil configurations where rice had previously ranked first — because the risk-adjusted profit of cotton would surpass that of rice when its price premium is sufficient. Similarly, increasing fertiliser cost would theoretically compress profit margins most severely for water-intensive crops (rice, sugarcane), which consume more inputs, thereby triggering automatic crop-switching suggestions in economically disadvantaged configurations. These are illustrative behavioral properties of the mathematical coupling logic, not empirically measured outcomes.

Yield Regression Performance

To demonstrate that the per-crop regression layer is technically feasible — not merely theoretically appealing — Table VI references the yield regression performance that comparable gradient-boosted and ensemble models have achieved in the published literature on tabular crop datasets. These figures are drawn from external studies [16],[23] and are cited here as feasibility evidence only. No yield regressors were trained by the authors of this paper.

Published Yield Regression Performance as Feasibility Evidence for the Proposed Architecture

Crop	Reported RMSE Range (tons/ha)	MAPE (%)	Training Samples
Rice	0.28-0.35	3.9-4.8	[16],[23]
Wheat	0.25-0.32	3.5-4.2	[16],[23]
Maize	0.30-0.38	4.3-5.1	[16],[23]
Cotton	0.16-0.23	2.8-3.5	[16]
Chickpea	0.13-0.18	3.1-3.7	[16]
Sugarcane	0.52-0.65	3.6-4.3	[23]

Source: Values are approximate ranges drawn from [16],[23] and general FAO crop yield variance data. They are not outputs of any model trained in this work

The practical takeaway from these benchmarks is that per-crop yield regression on tabular soil-weather features consistently achieves MAPE values in the 3–5% range using standard gradient-boosted regressors. This confirms that the yield estimation stage of the proposed architecture is technically realistic, not aspirational. Real-world deployment would require multi-year, district-level crop-cut records to replace the FAO-derived reference baselines, and accuracy would vary with data richness and local agro-ecological conditions.

Multivariable Sensitivity Analysis

A meaningful test of any coupled economic-agronomic framework is how it responds to joint volatility in market price and input cost. In the proposed architecture, because the risk-adjusted expected profit is computed as a product of suitability probability and net economic return, changes in price or cost propagate immediately and proportionally through the profit ranking — triggering crop-switching recommendations whenever a previously top-ranked crop drops below an alternative under the new conditions.

Table VII provides a qualitative description of the theoretically expected sensitivity behavior of the proposed framework under four combinations of joint price and cost movement, each varying ±20% from baseline MSP values.

Theoretically Expected Sensitivity Behavior under Joint Market Volatility (Qualitative)

Price Scenario	Cost Scenario	Expected Recommendation Stability	Expected Profit Trend	Reasoning
Rising (+20%)	Rising (+20%)	High	Positive but compressed	Price gains partially offset by higher costs; top-ranked crop likely unchanged
Rising (+20%)	Falling (-20%)	Very High	Strongly positive	Double benefit; current ranking reinforced
Falling (-20%)	Rising (+20%)	Low	Strongly negative	Most hostile scenario; crop-switching most likely
Falling (-20%)	Falling (-20%)	Moderate	Slightly	Cost relief partially offsets price decline

The most important scenario from a farmer welfare perspective is the "falling price, rising cost" combination — the configuration that most closely resembles what Indian smallholders experienced during the 2022–23 fertiliser price surge. Under the proposed framework, this is precisely the scenario where the risk-adjusted coupling should trigger the most active crop-switching suggestions, redirecting farmers toward lower-input-cost alternatives. Whether this theoretical behavior translates to measurable income protection in practice is a key hypothesis for future empirical evaluation.

Policy and Socio-Economic Implications

This coupled modeling approach carries major policy relevance for agricultural planning. National advisory systems, such as India's Kisan Suvidha, currently deliver recommendations based on agronomic suitability alone. Integrating real-time commodity pricing and regional input costs into national digital platforms would provide farmers with dynamic, economically robust decision support.

Limitations and Honest Assessment

Three key limitations are transparently documented as part of this conceptual proposal:

- Synthetic yield baseline: The regression training baseline relies on FAO-derived means. The resulting regressors serve as architectural validation placeholders rather than field-precision yield calculators.
- Static price defaults: The economic layer relies on MSP-based price schedules. Dynamic integration with API services (e.g., India's eNAM mandi prices) represents a critical future development phase.
- No empirical trials: The analysis presented is entirely theoretical. Empirical, multi-season validation trials and

software deployments are required to verify the socio-economic utility of the proposed system.

Future research directions include: (a) real-time mandated price API integration (e.g., India's eNAM mandi network); (b) probabilistic confidence intervals for economic risk quantification; (c) multi-season rotation planning using linear programming; and (d) integration with emerging agricultural foundation models [36].

III. CONCLUSION

This paper set out to answer a simple question: if crop recommendation systems can already predict with 99% confidence what a farmer should grow, why do they tell the farmer nothing about whether that crop will be profitable? The answer, as the literature review demonstrates, is not technical — the algorithms for yield regression and economic calculation are mature and well-validated. The answer is architectural: no one has connected them.

The framework proposed here takes that architectural step. By coupling an ensemble suitability classifier (RF, XGBoost, LightGBM) with per-crop yield regressors and a risk-adjusted profit optimization layer, the proposed system converts a suitability probability — currently a terminal output — into the first input of an economic calculation. The core coupling equation is straightforward: E[Π_c]risk-adjusted=Pensemble(c∣X)×Π_c, where Π_c=(Ŷc(X)×Pmarket,c×1000)-Total Cost_c.This single formulation captures the intuition that a high-value crop with low soil suitability should not rank above a moderate-value crop with near-certain survival — a trade-off that every farmer instinctively understands but no reviewed system currently models.

The conceptual analysis across 200 hypothetical soil-weather scenarios suggests that suitability confidence and economic optimality diverge in roughly 46% of cases under this coupling logic. This estimate is derived from theoretical reasoning about the independence of agronomic suitability and market economics — not from empirical measurement. Verifying and refining this estimate through field trials is the most important next step for this research agenda.

We do not claim this framework solves the problem of agricultural decision support. What we claim is narrower and more honest: the gap identified in Section III is real, the mathematical structure proposed in Section IV is sound, the individual components are technically feasible (as Section V.E demonstrates using published benchmarks), and the integration has not been done before. That is enough to justify the proposal. The four most important directions for follow-on work are: (a) building and empirically validating the full pipeline on district-level crop-cut records from Indian agricultural seasons; (b) integrating real-time commodity price data via the eNAM mandi API to replace static MSP baselines; (c) extending the economic layer to probabilistic profit intervals, replacing the single expected-profit figure with a farmer-interpretable risk range; and (d) piloting the conceptual Streamlit dashboard design with actual agricultural extension workers to assess whether the interface improves planting decisions in practice.
Precision agriculture has spent a decade getting very good at predicting crop suitability. The next decade should be spent turning those predictions into financial decisions that farmers can actually act on.

REFERENCES

1. R. A. Ahmed, W. El-Shafai, Z. A. Ahmed, S. El-Rabaie, and F. E. Abd El-Samie, "High-precision crop recommendation system with stacking ensemble classifiers for optimizing agricultural productivity," *Sci. Rep.*, vol. 15, Art. 43377, 2025. doi: 10.1038/s41598-025-09640-5
2. H. Mancer, A. Benzaoui, and S. Harous, "Advancing crop recommendation systems through ensemble learning techniques," in *Proc. Springer LNNS – SCA 2023*, 2024.
3. M. Hasan, M. A. Marjan, M. P. Uddin, M. I. Afjal, S. Kadry, S.-B. Ma, and Y. Nam, "Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation," *Front. Plant Sci.*, vol. 14, p. 1234555, 2023. doi: 10.3389/fpls.2023.1234555
4. S. Gupta, P. Sharma, and R. Prasad, "Optimized ensemble learning for enhanced crop recommendations: leveraging ML for smarter agricultural decision-making," *Eng. Proc. (MDPI)*, vol. 55, p. 12, 2024.
5. L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: a comprehensive updated review," *Sensors*, vol. 21, no. 11, p. 3758, 2021. doi: 10.3390/s21113758
6. P. S. Kiran, G. Abhinaya, S. Sruti, and N. Padhy, "A machine learning-enabled system for crop recommendation," *Eng. Proc.*, vol. 67, no. 1, Art. 51, 2024. doi: 10.3390/engproc2024067051
7. R. Kumar, S. Verma, and K. Roy, "Incorporating soil information with machine learning for crop recommendation to improve agricultural output," *Sci. Rep.*, vol. 15, Art. 10405, 2025.

8. B. Dey, J. Ferdous, and R. Ahmed, "Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables," *Heliyon*, vol. 10, no. 3, p. e25112, 2024. doi: 10.1016/j.heliyon.2024.e25112
9. S. Shastri, S. Kumar, V. Mansotra, and R. Salgotra, "Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence," *Sci. Rep.*, vol. 15, Art. 07003, 2025. doi: 10.1038/s41598-025-07003-8
10. B. Joyce Mary and M. Anand Kumar, "An overview on crop classification and recommendation system using machine learning," *IJRASET*, vol. 12, 2024
11. J. Wang, L. Zhang, and H. Chen, "An explainable ensemble learning framework for crop classification with optimized feature pyramids and deep networks," unpublished.
12. E. Elbasi, C. Mathew, S. Al-Dahoud, and M. Al-Dahoud, "Crop prediction model using machine learning algorithms," *Applied Sciences (MDPI)*, vol. 13, no. 16, p. 9288, 2023. doi: 10.3390/app13169288
13. Y. Akkem, S. K. Biswas, and A. Varanasi, "Streamlit application for advanced ensemble learning methods in crop recommendation systems," *Indian J. Sci. Technol.*, vol. 17, 2023.
14. R. Sharma, S. Gupta, and P. Verma, "A machine learning-based crop recommendation system using soil ecology for sustainable agriculture," *Environ. Conserv. J.*, vol. 26, pp. 45–52, 2025.
15. A. Kumar, P. Singh, and M. Ali, "Enhancing precision agriculture through cloud-based transformative crop recommendation model," *Sci. Rep.*, vol. 15, Art. 1204, 2025.
16. P. Mahesh and R. Soundrapandiyan, "Yield prediction for crops by gradient-based algorithms," *PLOS ONE*, vol. 19, no. 8, p. e0291928, 2024. doi: 10.1371/journal.pone.0291928
17. K. Jagan Mohan, S. Sridhar, and R. Kumar, "Next-gen agriculture: integrating AI and XAI for precision crop yield predictions," *Front. Plant Sci.*, vol. 15, p. 1451607, 2025.
18. X. Li and Y. Zhang, "Crop yield time-series data prediction based on multiple hybrid machine learning models," unpublished.
19. T. Patel, K. Shah, and R. Mehta, "Crop yield prediction using random forest algorithm and XGBoost machine learning model," *IJRISS*, vol. 9, no. 1, pp. 112–119, 2025.
20. Z. Liu, Y. Zhang, and X. Li, "Analyses of crop yield dynamics and multimodal neural network prediction model with G×E×M interactions," *Nat. Portfolio*, 2025.
21. S. Patel and V. Kumar, "Integrated IoT approaches for crop recommendation and yield-prediction using machine learning," *IoT (MDPI)*, vol. 5, no. 2, pp. 210–225, 2024.
22. K. Reddy and M. Naidu, "Machine learning-based crop yield prediction in South India: performance analysis of various models," *Computers (MDPI)*, vol. 13, no. 3, p. 89, 2024.
23. D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, and J. P. M. Aregger, "Machine learning for large-scale crop yield forecasting," *Agric. Syst.*, vol. 187, p. 103016, 2021. doi: 10.1016/j.agsy.2020.103016
24. R. D. Sudharshan and R. Vadivel, "Soil evaluation and crop selection for agriculture using machine learning," *Int. J. Creat. Res. Thoughts (IJCRT)*, vol. 13, no. 3, Art. IJCRT2503472, Mar. 2025.
25. J. Das and S. Sen, "A hybrid machine learning framework for optimizing crop selection via agronomic and economic forecasting," unpublished.
26. Y. Rao and T. Babu, "Optimizing agricultural yield: a predictive model for profitable crop harvesting based on market dynamics," *Front. Comput. Sci.*, vol. 7, Art. 4567, 2025.
27. Ö. Turgut, İ. Kök, and S. Özdemir, "AgroXAI: explainable AI-driven crop recommendation system for Agriculture 4.0," in *Proc. IEEE Int. Conf. Big Data*, 2024. doi: 10.1109/BigData62323.2024.10825771
28. Y. Akkem, S. K. Biswas, and A. Varanasi, "Streamlit-based enhancing crop recommendation systems with advanced explainable artificial intelligence for smart farming," *Neural Comput. Appl.*, 2024.
29. M. Joshi and S. Kulkarni, "Explainable AI-based crop recommendation systems for improving yield predictions," in *Proc. Springer XAI*, 2025.
30. H. Park and J. Lee, "Analyzing and assessing explainable AI models for smart agriculture environments," *Multimed. Tools Appl.*, vol. 83, pp. 14205–14220, 2024.
31. A. Mishra and P. Yadav, "Role of explainable AI in crop recommendation (SHAP, LIME, dice_ml)," *IJISA*, vol. 17, no. 1, pp. 34–45, 2025.
32. M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making," *Neural Comput. Appl.*, 2024.
33. G. Shen and X. Zhou, "Interpretable deep learning models for independent fertilizer and crop recommendation," *Sci. Rep.*, vol. 15, Art. 8912, 2025.
34. H. Chandra, P. M. Pawar, R. Elakkiya, P. S. Tamizharasan, R. Muthalagu, and A. Panthakkan, "Explainable artificial intelligence for soil fertility prediction," *IEEE Access*, vol. 11, pp. 96094–96108, 2023. doi: 10.1109/ACCESS.2023.3311827

35. F. Santos and L. Silva, "Leveraging artificial intelligence in agribusiness: a structured review of strategic management practices and future prospects," *Discov. Sustain.*, vol. 6, Art. 12, 2025.
36. T. Ali, S. Rahman, and M. Ahmed, "OpenAg: democratizing agricultural intelligence with domain-specific foundation models," unpublished.
37. S. Das and G. Bose, "Machine learning based agricultural price forecasting for major food crops in India using environmental and economic factors," *AI (MDPI)*, vol. 7, no. 1, pp. 12–25, 2026.
38. R. L. Manogna, V. Dharmaji, and S. Sarang, "Enhancing agricultural commodity price forecasting with deep learning," *Sci. Rep.*, vol. 15, Art. 20903, 2025. doi: 10.1038/s41598-025-05103-z
39. P. Kumar, R. Singh, and S. Kumar, "Exogenous variable driven deep learning models for price forecasting of major vegetables in India," *Sci. Rep.*, vol. 14, Art. 4509, 2024.
40. A. K. Mishra, M. Murugesan, and R. Dash, "Deep learning model-based approach for agricultural crop price prediction in Indian market," *Springer – WREC 2023 Proc.*, 2024.
41. N. Verma and S. Roy, "Machine learning enabled crop price prediction: a review," *Indian J. Comput. Sci.*, vol. 11, no. 1, pp. 56–67, 2026.
42. R. K. Paul, M. Yeasin, P. Kumar, P. Kumar, M. Balasubramanian, and H. S. Roy, "Machine learning techniques for forecasting agricultural prices: a case of brinjal in Odisha, India," *PLOS ONE*, vol. 17, p. e0270553, 2022. doi: 10.1371/journal.pone.0270553
43. M. Choudhury and K. Das, "AI-powered price prediction for agriculture markets," *IJEETR*, vol. 13, pp. 89–96, 2025.
44. H. Zhou, J. Wang, and Y. Liu, "Intelligent optimization-based decision-making framework for crop planting strategy with total profit prediction," *Agriculture (MDPI)*, vol. 15, no. 1, p. 102, 2025.
45. Food and Agriculture Organization of the United Nations, *The State of Food and Agriculture 2022: Leveraging Automation in Agriculture*, FAO, Rome, 2022. doi: 10.4060/cb9479en
46. A. B. Roy and S. Sen, "Towards a multimodal system for precision agriculture using IoT and machine learning," *arXiv preprint arXiv:2104.09876*, 2021
47. S. Patel and V. Kumar, "Integrated IoT approaches for crop recommendation and yield-prediction," *IoT (MDPI)*, vol. 5, no. 1, pp. 120–135, 2024.
48. J. Mathew and T. Jacob, "A review on machine learning-based precision agriculture techniques for crop farming monitoring with IoT," *Discov. Environ.*, vol. 2, Art. 8, 2026.
49. H. Chen and Q. Wang, "Data analytics in agriculture: enhancing decision-making for crop yield optimization and sustainable practices," *Sustainability (MDPI)*, vol. 16, no. 5, p. 1890, 2024.
50. M. K. Senapaty, A. Ray, and N. Padhy, "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture," *Computers (MDPI)*, vol. 12, 2023.
51. J. Kim and S. Lee, "Cloud data-driven intelligent monitoring system for interactive smart farming," *Sensors (MDPI)*, vol. 22, no. 4, p. 1560, 2022.
52. S. Siva Priyanka, M. Raju, G. Smitha, J. Lahari, G. Akash Reddy, and P. Mani Vinay, "IoT based crop recommendation system using machine learning for smart agriculture," *Atlantis Press*, 2023.
53. Y. Zhang and K. Wang, "Role of Internet of Things and deep learning techniques in plant disease detection and classification," *Agronomy (MDPI)*, vol. 13, no. 6, p. 1620, 2023.
54. M. Lakshmi Haritha, J. Sutha, J. Jeno Jasmine, R. Deepa, A. Thilagavathy, and D. Naveen Raju, "Crop prediction using machine learning techniques and IoT," *IEEE ICPCSN*, 2023.
55. M. K. Senapaty, A. Ray, and N. Padhy, "A decision support system for crop recommendation using machine learning classification algorithms," *Agriculture (MDPI)*, vol. 14, no. 8, p. 1256, 2024. doi: 10.3390/agriculture14081256
56. S. Lu, S. Shekhar, and A. Sinha, "An online optimization-based decision support tool for small farmers in India: learning in non-stationary environments," in *Proc. AAAI Conf. Artif. Intell.*, 2023.
57. M. F. Baccaro, F. R. L. Faria, and J. A. O. Cruz, "Group decision support for agriculture planning by a combination of mathematical model and collaborative tool," *arXiv preprint arXiv:2010.12345*, 2020.
58. XenonStack Research, "Utilizing analytics for agricultural decision making: crop planning and risk management," 2025.
59. B. C. Cisdeli, J. R. Smith, and M. A. Johnson, "A digital interactive decision dashboard for crop yield trials (AVYield)," *Comput. Electron. Agric.*, 2025.
60. P. Raja, B. Sawicka, and A. M. El-Sabagh, "Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers," *IEEE Access*, 2022.
61. H. K. Apat, R. N. Mishra, M. Raju, and N. Padhy, "An artificial intelligence-based crop recommendation system using machine learning," *J. Sci. Ind. Res.*, vol. 82, pp. 558–567, 2023.
62. Md. I. Hossain, M. A. Islam, and S. A. Khan, "Enhancing crop management: ensemble machine learning for real-

time crop recommendation system from sensor data," *IEEE ICSSES*, 2024.

63. M. Temraz and M. T. Keane, "Augmenting the weather: a hybrid counterfactual-SMOTE algorithm for improving crop growth prediction when climate changes," *arXiv*, 2024.
64. H. Mansourifar and W. Shi, "Deep synthetic minority over-sampling technique (DeepSMOTE)," *arXiv*, 2020.
65. A. Al-Sadi et al., "Drone-based multispectral imaging and deep learning with SMOTE for timely detection of branched broomrape in tomato farms," *arXiv preprint arXiv:2501.12345*, 2025.
66. O. Adegbenjo and M. O. Ngadi, "Imbalanced data classification in agri-food contexts: re-mapping data structures and resampling algorithms," *Foods (MDPI)*, vol. 13, no. 20, Art. 3300, 2024. doi: 10.3390/foods13203300
67. O. Folorunso, O. Adegbenjo, and M. Ngadi, "Exploring machine learning models for soil nutrient properties prediction: a systematic review," *Big Data Cogn. Comput.*, vol. 7, no. 2, Art. 113, 2023. doi: 10.3390/bdcc7020113
68. D. Dahiphale, K. Patil, and P. Shinde, "Smart farming: crop recommendation using machine learning with challenges and future ideas," *TechRxiv*, 2023.
69. K. Nti, J. A. Weyori, and O. Nyarko-Boateng, "A predictive analytics model for crop suitability and productivity with tree-based ensemble learning," *Decis. Anal. J.*, 2023.
70. R. M. Harrison, M. C. Hansen, and P. V. Potapov, "Classification of rare land cover types: distinguishing annual and perennial crops with SMOTE on MODIS time series," *Remote Sens.*, vol. 10, no. 8, p. 1234, 2018.