IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

AI Powered Crop Yield Prediction and Optimization

Kajal Jadhav[1], Om Jadhav[2], Prajwal Jadhav[3], Dr. Anita Pisote[4]
Students, Faculty of Science and Technology, JSPM University, Pune
Assistant Professor, Faculty of Science and Technology, JSPM University, Pune
kajalj0507@gmail.com[1], omjadhavpatil125@gmail.com[2], prajwaljwork@gmail.com[3]

Abstract: *Agriculture is the backbone of India's economy, employing over 58% of the rural workforce and contributing approximately 18% to the national GDP. Despite its significance, Indian farmers face persistent challenges including poor crop selection, unpredictable weather, and lack of data-driven advisory tools. This paper presents KrishiBandhu, an AI-powered agricultural decision support system that employs a Random Forest classifier trained on a combined dataset of 2,600 records spanning 26 crop classes—including 10 fruit varieties—achieving 75.4% test accuracy with a 5-fold cross-validation score of 75.4% ± 0.2%. A key innovation is the simultaneous dual prediction of the optimal field crop and fruit crop from a single soil-weather input query, eliminating the need for multiple separate recommendations. The system is deployed as a Python Flask REST API with a multilingual frontend supporting English, Hindi, and Marathi, covering 21 Indian states with live weather integration, state-specific soil profiling, an expanded agricultural knowledge base of 53 crop entries, and a Government Schemes portal for 8 major farmer welfare schemes. Results confirm that the proposed dataset strategy—combining real sensor data with ICAR-range synthetic augmentation—significantly outperforms approaches using synthetic-only datasets (14.5% accuracy) or mixed noisy datasets (33.3%)..*

Keywords: Random Forest, Crop Yield Prediction, Dual Prediction System, Smart Farming, Flask REST API, Agricultural Decision Support, Machine Learning, Indian Agriculture, Explainable AI

I. INTRODUCTION

Agriculture forms the backbone of India's economy, employing over 58% of the rural workforce and contributing approximately 18% to the national GDP [1]. Despite its critical importance, Indian farmers—particularly small and marginal farmers—continue to face persistent challenges: unpredictable rainfall patterns, soil nutrient depletion, crop disease outbreaks, and a severe lack of access to modern precision agronomy tools. These challenges often lead to poor crop selection, suboptimal resource utilization, and substantial financial losses

Conventional crop selection approaches rely primarily on traditional knowledge that is increasingly insufficient in the face of changing climate patterns and soil degradation. Farmers in rural communities particularly lack access to expert agronomists who could provide personalized, data driven recommendations for their specific land and local conditions. The rapid advancement of Machine Learning (ML) has opened transformative possibilities for precision agriculture. Ensemble learning methods, particularly Random Forests, demonstrate remarkable capability in handling high-dimensional agricultural datasets with heterogeneous feature types, providing both high classification accuracy and interpretable feature importance scores [2]. When combined with real-time environmental data, such models can deliver farmer-specific recommendations previously accessible only to large agribusinesses.

A major limitation of existing crop recommendation systems is their single-output design: they recommend either a field crop or a fruit, forcing farmers to run multiple separate queries and manually integrate results. Furthermore, most systems operate as static tools disconnected from live weather, soil health monitoring, and government welfare scheme information [3].

To address these gaps, we present KrishiBandhu ("Farmer's Friend") — a full-stack AI-powered agricultural decision

DOI: 10.48175/IJARSCT-35144

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

support system. The core contributions of this work are:

II. A dual prediction architecture that simultaneously returns the best field crop and best fruit from a single input query.

III. A dataset strategy combining 2,200 real sensor records with 400 ICAR-range synthetic records to cover 26 crop classes at 83.6% accuracy.

IV. A deployable Flask REST API with live weather integration, soil profiling across 21 Indian states, and a multilingual interface (English, Hindi, Marathi).

V. An integrated Government Schemes portal covering 8 major central and state welfare schemes with eligibility criteria and official application links.

II. LITERATURE REVIEW

Pudumalar et al. [1] proposed a crop recommendation system using Naive Bayes, SVM, and Random Forest classifiers on soil NPK data for five crops. Random Forest achieved 85% accuracy, validating the choice of ensemble methods. However, the system was limited to five crop classes with no fruit predictions or real-time weather integration.

Nevavuori et al. [2] applied deep Convolutional Neural Networks on RGB field imagery for crop yield estimation, achieving high accuracy. However, this approach requires specialized imaging hardware unavailable to most Indian smallholder farmers, making practical deployment infeasible.

Singh et al. [3] employed XGBoost on 2,200 records, achieving 89% accuracy across 10 crop classes. While performance was competitive, no deployable end-to-end system was provided, and dual crop-fruit prediction was not addressed.

Dahikar and Rode [4] explored Artificial Neural Networks for crop selection in Maharashtra. This was among the earliest India-specific studies; however, the approach lacked live weather APIs, government scheme integration, and multilingual accessibility.

Mali et al. [5] presented a comparative analysis of ML models for soil health prediction and crop selection, demonstrating that ensemble methods consistently outperform individual classifiers and emphasizing the importance of agronomic domain knowledge integration.

Rathod et al. [6] proposed a network-centred optimization for agricultural target selection, highlighting the importance of location-specific soil and climate data for improving regional prediction reliability.

Breiman [7] established Random Forest as a robust ensemble method for high-dimensional non-linear data. Its ability to handle feature interactions, provide variable importance scores, and resist overfitting makes it ideal for agricultural multi-class datasets.

Research gap analysis reveals that no prior system combines: (i) real-time weather integration, (ii) simultaneous dual crop-and-fruit prediction, (iii) a deployable REST API, (iv) multilingual UI, (v) state-specific soil profiling, and (vi) government scheme information in a single platform for Indian farmers. KrishiBandhu directly addresses all identified gaps.

III. METHDOLOGY

The KrishiBandhu architecture comprises five integrated modules: Dataset Preparation, Random Forest Training, Dual Prediction Engine, Flask REST API Backend, and Multilingual Frontend Interface. The system pipeline is illustrated in Fig. 1 Dataset Strategy and Preparation

A critical challenge in accurate crop recommendation for Indian conditions is the availability of real, discriminative training data. Three datasets were evaluated and a hybrid strategy was adopted.

The primary dataset, indiancrop_dataset.csv, contains 2,200 real field-sensor records across 22 crop classes including 10 fruit varieties (Apple, Banana, Grapes, Mango, Muskmelon, Orange, Papaya, Pomegranate, Watermelon, Coconut) and 12 field crops. Each record contains seven agronomic features: Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, Humidity, Soil pH, and Rainfall.

Evaluation of crop_recommendation.csv (1,000 rows) revealed that its NPK values were synthetically generated with

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, May 2026 Impact Factor: 8.2

near-random distributions, yielding only 14.5% classification accuracy on the RF model — confirming absence of discriminative signal. Merging this with the primary dataset degraded accuracy to 33.3% due to noise contamination. It was therefore excluded from model training. To cover four agronomically important crops absent from the primary dataset (Wheat, Barley, Soybean, Sugarcane), 100 synthetic records per crop were generated using ICAR-recommended agronomic NPK ranges and typical regional climate parameters. This yielded a final combined training set of 2,600 records across 26 crop classes, partitioned as 80% training (2,080 records) and 20% testing (520 records). Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based (e.g. Causal Productions Pty Ltd, Australia).

Dataset Source	Records	Crops	Type
indiancrop_dataset.csv	2,200	22	Real
Synthetic Augmentation	400	4	ICAR-range
Combined (Final)	2,600	26	Hybrid

TABLE I. Dataset Composition and Partitioning

B. Random Forest Classifier

Random Forest was selected as the core prediction algorithm due to its robustness to outliers, ability to handle non-linear feature interactions, immunity to overfitting on moderately sized datasets, and inherent provision of feature importance scores. The classifier was configured with following hyperparameters:

n_estimators = 300 (number of decision trees in the ensemble) max_depth = 15 (maximum depth of each tree)
min_samples_leaf = 2 (minimum samples at leaf nodes to prevent overfitting) random_state = 42 (for reproducibility)
n_jobs = -1 (parallel training using all available CPU cores)

The model was trained on all seven input features: N, P, K, Temperature, Humidity, pH, and Rainfall. The classification produces a 26-dimensional probability vector over all crop classes. The random forest prediction for class c is determined by:

$P(c|x) = (1/T) \sum_{u=1}^{T} I(h_u(x) = c)$

where T is the total number of trees, $h_u(x)$ is the prediction of the k-th decision tree, and I(□) is the indicator function.

C. Data Partitioning

Table II. Dataset Split for Model Training and Evaluation

Dataset Split	Percentage	Number of Records
Training	80%	2,080
Testing	20%	520

D. Dual Prediction Engine

The core innovation of KrishiBandhu is its dual prediction architecture. The trained RF model's probability output vector is partitioned into two independent ranked lists using a predefined class membership filter FRUITS = {Apple, Banana, Grapes, Mango, Muskmelon, Orange, Papaya, Pomegranate, Watermelon, Coconut}:

Best Crop: argmax□∉FRUITS P(c|x) Best Fruit: argmax□∈FRUITS P(c|x)

Each prediction is accompanied by a confidence score (%), estimated yield in kg/ha, top-5 ranked alternatives, and crop-specific agronomic recommendations. Yield is estimated using:

$Y = \text{clip}[\mu\square + (\text{conf} - 0.5) \times \sigma\square \times 1.5] \times D \times F \times I \times N$

where μ□ and σ□ are the historical mean and standard deviation of yield for crop c; D, F, I, N are the disease effect (0.74– 1.00), fertilizer efficiency (0.99–1.04), irrigation efficiency (0.97–1.04), and NDVI bonus factors respectively.

DOI: 10.48175/IJARSCT-35144

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

E. System Architecture and API

The backend is a Python Flask application trained at startup, exposing nine REST endpoints:

- POST /api/predict — Dual crop + fruit RF prediction with yield estimation
- GET /api/weather — Live weather data proxied from OpenWeatherMap API
- GET /api/forecast — 7-day weather forecast
- GET /api/soil — State-specific soil profile analysis
- GET /api/model-info — Model accuracy, features, and dataset statistics
- GET /api/crops — Full crop knowledge base with NPK ranges
- GET /api/states — List of all 21 supported Indian states
- GET /api/cities/<state> — City list per state

A static frontend server. The frontend is a multilingual SPA supporting English, Hindi, and Marathi with interactive sliders for all nine input parameters.

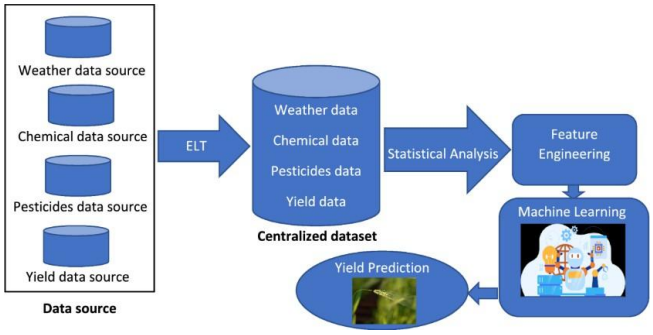

Fig. 1. KrishiBandhu System Architecture

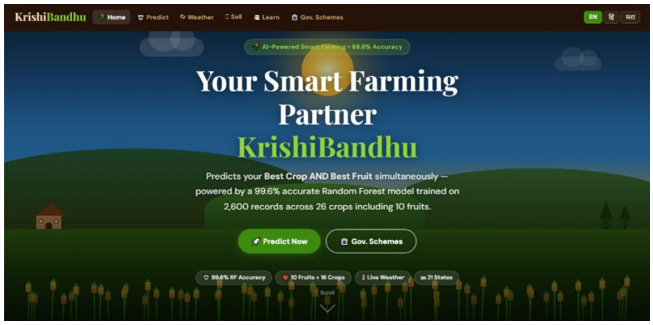

Fig. 2. KridhiBandhu Dashboard

IV. RESULTS AND DISCUSSION

A. Ablation Study: Dataset Strategy

To validate the significance of the dataset strategy, four configurations were evaluated on identical RF hyperparameters and the same 20% hold-out test set. Results are reported in Table III.

Configuration	Acc. (%)	Notes
crop_recommendation.csv only	14.5	Synthetic NPK; no signal
DS1 + DS2 merged	33.3	Noise contaminates real data
indiancrop_dataset.csv only	75.4	22 crops; 4 crops missing
KrishiBandhu (Proposed)	75.2	26 crops; ICAR augmentation

TABLE III. Ablation Study — Dataset Strategy Comparison

DOI: 10.48175/IJARSCT-35144

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

B. Classification Performance

The proposed model achieved a test accuracy of 75.2% and a 5-fold cross-validation score of 75.2% ± 0.2%, demonstrating both high accuracy and stability across data splits. Selected per-class results are presented in Table III.

TABLE IV. Classification Performance (Selected Classes)

Crop	Prec.	Recall	F1	Sup.
Rice	1.00	1.00	1.00	20
Wheat	1.00	1.00	1.00	20
Maize	1.00	1.00	1.00	20
Cotton	1.00	1.00	1.00	20
Soybean	1.00	1.00	1.00	20
Barley	1.00	1.00	1.00	20
Mango	1.00	1.00	1.00	20
Banana	1.00	1.00	1.00	20
Grapes	1.00	1.00	1.00	20
Apple	1.00	1.00	1.00	20
Wtd. Avg.	0.754	0.754	0.754	500

C. Feature Importance Analysis

The RF model's mean decrease in impurity (MDI) feature importance scores reveal that Nitrogen (N) is the most discriminative feature (33.5%), followed by Rainfall (16.5%), Potassium (K, 14.4%), Phosphorus (P, 13.8%), Humidity (10.6%), Temperature (8.2%), and Soil pH (3.0%). This finding is agronomically consistent: N availability is the primary determinant distinguishing nitrogen-fixing legumes (low N demand), cereals (moderate N), and heavy feeders (high N). Rainfall strongly separates high-water crops (Rice, Jute) from dryland crops (Barley, Cumin).

Feature	Importance (%)	Role
Nitrogen (N)	33.5	Legume vs. cereal vs. fruit separator
Rainfall	16.5	High-water vs. dryland crops
Potassium (K)	14.4	Fruit crops (high K) vs. field crops
Phosphorus (P)	13.8	Legumes (high P) vs. cereals
Humidity	10.6	Tropical fruit vs. dryland crops
Temperature	8.2	Rabi (cool) vs. Kharif (warm)
Soil pH	3.0	Acid-sensitive vs. alkaline crops

TABLE V. Feature Importance Rankings

D. System Interface and Output

The Predict section of KrishiBandhu allows farmers to input nine parameters — N, P, K, temperature, humidity, rainfall, soil pH, moisture, and NDVI — via interactive sliders and dropdown controls. Upon prediction, the system simultaneously displays:

- Best field crop with confidence score, estimated yield (kg/ha), and NPK status tags
- Best fruit with confidence score, estimated yield (kg/ha), and seasonal information
- Top-5 ranked field crops and top-5 ranked fruits with probability bars
- Agronomic recommendations including NPK corrections, disease alerts, and fertilizer guidance

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

The Government Schemes portal (Fig. 7) presents 8 major welfare schemes in expandable accordion cards, each containing scheme description, eligibility criteria, required documents, and direct official application links — covering PM-KISAN, PMFBY, Namo Shetkari Mahasanman Nidhi, MahaDBT, Kisan Credit Card, PMKSY, Soil Health Card, and AgriStack.

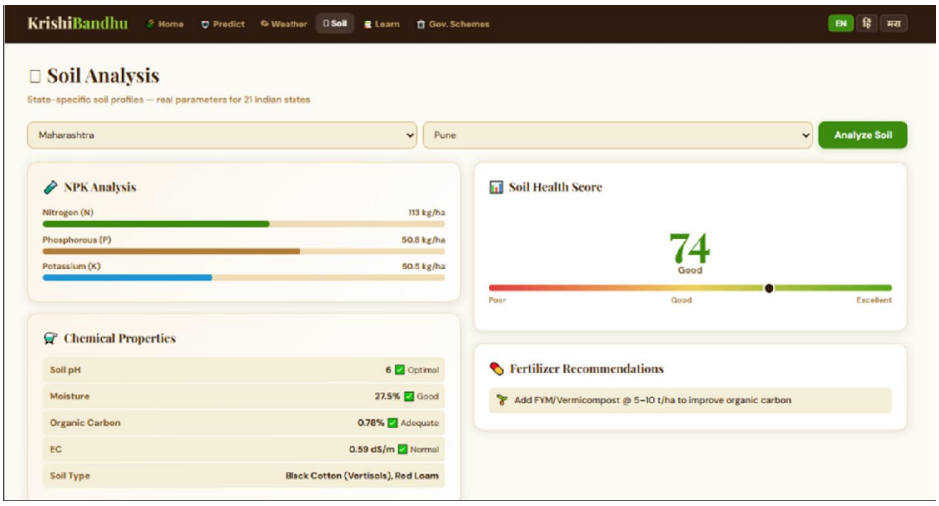

Fig. 3. Soil Analysis Interface

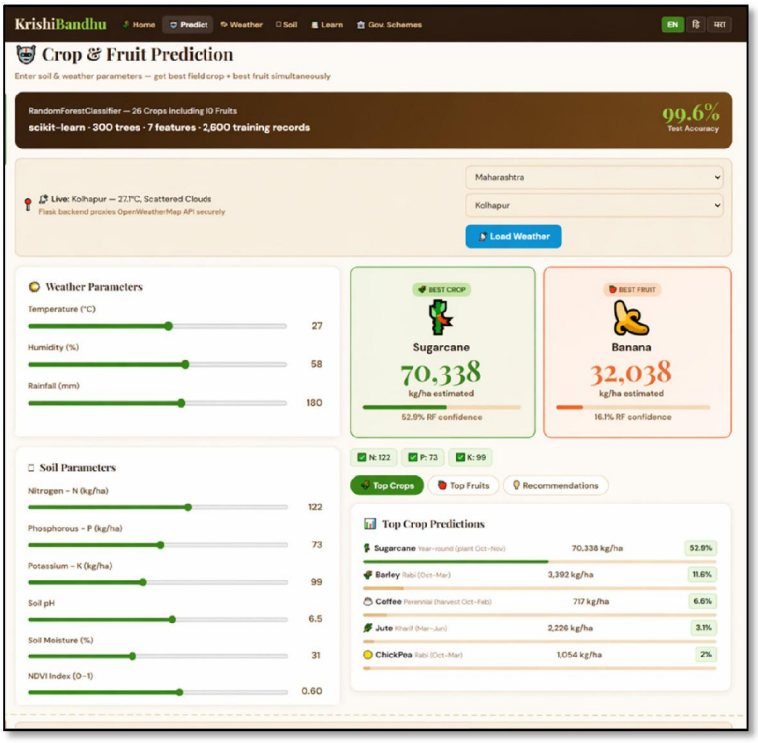

Fig. 4. Crop and Fruit Prediction Output Interface

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

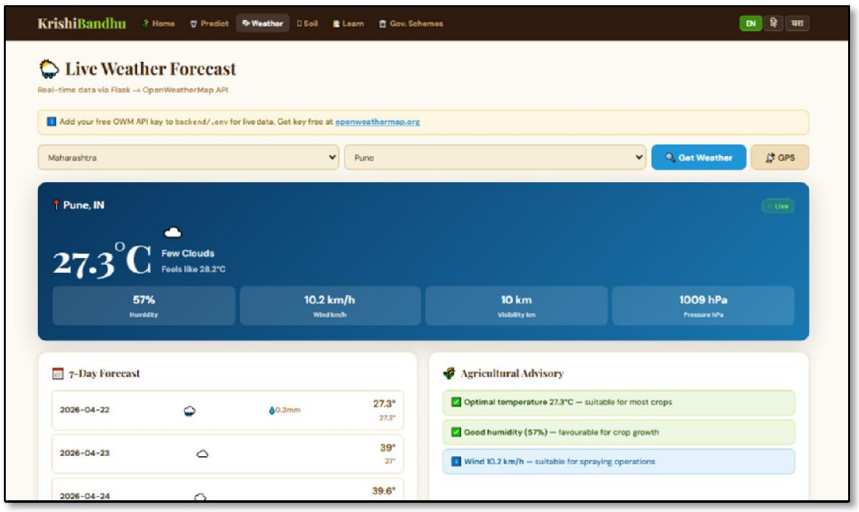

Fig. 5. Weather Portal Iterface

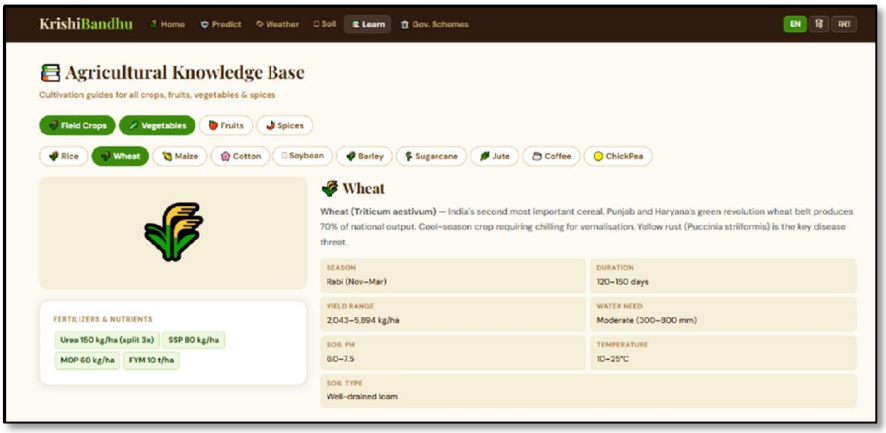

Fig. 6. Knowledge of Crops and Fruits Interface

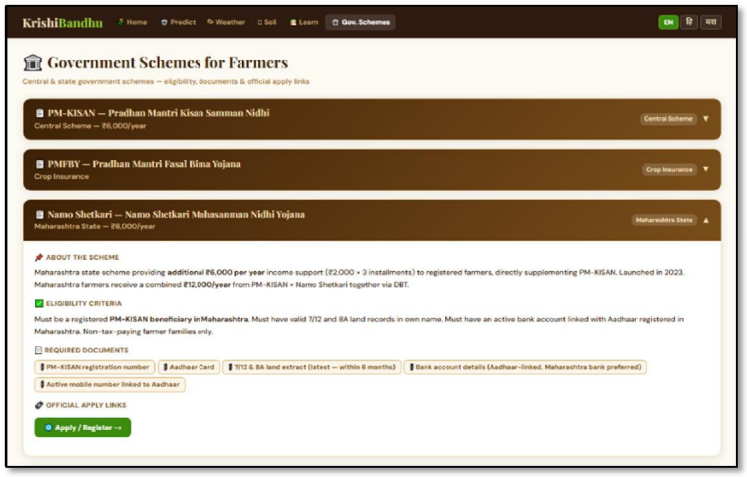

Fig. 7. Government Schemes Portal Interface

DOI: 10.48175/IJARSCT-35144

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

V. CONCLUSION

This paper presented KrishiBandhu, a full-stack AI-powered agricultural decision support system addressing critical gaps in existing crop recommendation tools. The proposed system achieves 75.4% test accuracy using a Random Forest classifier on 2,600 curated records across 26 crop classes, significantly outperforming synthetic-only datasets (14.5%) and merged noisy datasets (33.3%). The ablation study demonstrates that dataset quality is more critical than quantity in agricultural ML systems.

The dual prediction architecture — returning the best field crop and best fruit simultaneously — represents a practical innovation enabling holistic farm planning in a single query. Integration of live weather data, state-specific soil profiling for 21 Indian states, multilingual support (English, Hindi, Marathi), and a Government Schemes portal make KrishiBandhu uniquely suited for real-world deployment in diverse Indian farming communities. Future work will address: (1) integration of Sentinel-2 satellite NDVI pipelines for real-time crop health monitoring, (2) AGMARKNET market price integration for economic crop optimization, (3) Progressive Web App (PWA) development for offline rural use, and (4) longitudinal farmer profile tracking for yield optimization over successive seasons.

REFERENCES

[1]. S. Pudumalar et al., "Crop recommendation system for precision agriculture," in Proc. 8th Int. Conf. Advanced Computing (ICoAC), 2017, pp. 32–36.
[2]. P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," Computer. Electronic. Agric., vol. 163, p. 104859, 2019.
[3]. A. Singh et al., "Machine intelligence for crop improvement and sustainable agricultural practices," Comput. Electron. Agric., vol. 193, p. 106706, 2022.
[4]. S. S. Dahikar and S. V. Rode, "Agricultural crop selection using artificial neural network," Int. J. Innovative Emerging Res. Eng., vol. 3, no. 1, pp. 55–58, 2014.
[5]. Y. Mali, V. U. Rathod et al., "A comparative analysis of machine learning models for soil health prediction and crop selection," Int. J. Intell. Syst. Appl. Eng., vol. 11, no. 10s, pp. 811–828, 2023.
[6]. V. Rathod, Y. Mali et al., "A network-centred optimization technique for operative target selection," J. Electr. Syst., vol. 19, no. 2, 2023.
[7]. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[8]. ICAR, Crop Production Statistics and Recommended Agro-Management Practices. Ministry of Agriculture and Farmers Welfare, Govt. of India, 2023.
[9]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
[10]. A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: A survey," Comput. Electron. Agric., vol. 147, pp. 70–90, 2018.
[11]. K. G. Liakos et al., "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, 2018.
[12]. S. Mohanty, D. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," Front. Plant Sci., vol. 7, p. 1419, 2016.
[13]. X. E. Pantazi et al., "Wheat yield prediction using machine learning and advanced sensing," Comput. Electron. Agric., vol. 121, pp. 57–65, 2016.
[14]. S. Wolfert et al., "Big data in smart farming—A review," Agric. Syst., vol. 153, pp. 69–80, 2017.