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 2, November 2026
Impact Factor: 8.2

AI-Based Smart Agriculture: An Integrated Machine Learning Framework for Crop Recommendation, Fertilizer Recommendation and Crop Yield Prediction

Mr. Pol Ganesh Prakash, D. D. Patil, D. D. Patil
Department of MCA
Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, Maharashtra
gpol540@gmail.com

Abstract : *Agricultural productivity in India is strongly affected by weather variability, soil degradation and the over- or under-use of fertilizers. Farmers still tend to rely on personal experience when choosing a crop, estimating its yield and deciding how much fertilizer to apply. This paper presents an integrated, web-based decision-support system that performs three tasks within a single framework: crop recommendation, fertilizer recommendation and crop yield prediction. The system takes soil nutrients (nitrogen, phosphorus and potassium), temperature, humidity, soil pH, rainfall, soil moisture, soil type and crop type as inputs. The pipeline consists of data acquisition from public repositories, cleaning and normalisation, correlation-based feature selection, and model training using Artificial Neural Networks (ANN) and Random Forest (RF). Model performance is evaluated with cross-validation using accuracy, precision, recall, F1-score and correlation-based measures. In the reported experiments, the ANN model for yield prediction reached an accuracy of 94.70% and the Random Forest model for fertilizer recommendation reached 93.94%, and the crop recommendation model [insert accuracy] was the best performer for its task. The models are packaged in a Java (JSP/Servlet, Weka, MySQL) web application that lets farmers obtain all three recommendations from one input form..*

Keywords: smart agriculture, crop recommendation, crop yield prediction, fertilizer recommendation, artificial neural network, random forest, feature selection, precision agriculture.

I. INTRODUCTION

Agriculture underpins food security and rural livelihoods in India, yet it is exposed to erratic rainfall, changing temperatures, soil erosion and nutrient mismanagement. Excessive fertilizer use raises costs and harms soil health, while insufficient use lowers yields. Because soil, weather and crop variables interact in non-linear ways, traditional decision-making based on observation and experience often produces sub-optimal choices.

Machine learning (ML) can learn these relationships from data. Existing studies, however, usually address one question at a time: predicting yield, or recommending a crop, or recommending a fertilizer. Farmers need all three answers together, in a form they can use without technical knowledge.

This work makes the following contributions:

An integrated framework that combines crop recommendation, fertilizer recommendation and yield prediction in a single pipeline, where the recommended crop feeds the fertilizer and yield modules.

A preprocessing and feature-selection stage (missing-value replacement, min-max normalisation and correlation-based feature subset selection) applied before model training.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40704

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 2, November 2026
Impact Factor: 8.2

A comparison of ANN, Random Forest and Linear Regression based approaches for the three prediction tasks, evaluated with cross-validation.

A deployable web application built with Java, JSP/Servlets, Weka and MySQL that exposes the models to end users.

The rest of the paper is organised as follows. Section 2 reviews related work, Section 3 describes the datasets and methodology, Section 4 covers implementation, Section 5 presents results, Section 6 discusses limitations, and Section 7 concludes.

II. RELATED WORK

Yield prediction. Vinod et al. [1] apply regression techniques to environmental and soil data for yield forecasting. Reddy and Kumar [6] use supervised learning on climatic and soil parameters and report improved prediction precision. Gunjan et al. [5] compare Linear Regression, KNN and Random Forest on real agricultural datasets and find that ensemble methods outperform conventional approaches. Joshua et al. [9] evaluate several ML models across datasets and crops and conclude that advanced ensembles generalise best. Bali and Singla [4] survey neural networks, support vector machines and ensemble models for yield prediction, and Morales and Villalobos [8] study ML-based yield prediction under past and future climate scenarios. Agarwal and Tarar [7] combine Random Forest with a Convolutional Neural Network and show that a hybrid model captures non-linear dependencies better than single algorithms.

Crop recommendation. Shariff et al. [2] use Decision Trees and Random Forests to recommend crops from soil nutrients, weather and regional conditions. Dey et al. [11] recommend crops for India based on NPK, soil pH and three climatic variables. Okwii [13] and Gottemukkala et al. [15] build recommendation systems around NPK sensors, and Senapaty et al. [10] propose an IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture.

Fertilizer recommendation and integrated systems. Tanaka et al. [12] examine whether ML models (gradient boosting, neural networks, linear regression) can provide accurate fertilizer recommendations. Hegde and Kumar [3] combine yield and price forecasting in one application, and Sundaresan et al. [14] combine ML with IoT for smart farming.

Research gap. The reviewed studies share several weaknesses: they focus on specific crops or regions and adapt poorly to changing conditions [1], [3], [5], [6], [9]; nutrient-based recommenders often ignore climatic variation [2], [10], [11], [13], [15]; hybrid and deep models need large datasets and computing resources [7], [8]; IoT-based solutions face sensor calibration, network reliability and cost issues [10], [14]; deep models can overfit small datasets and are hard to interpret [4], [15]; and accuracy depends on high-quality, regularly updated soil data [12]. Few existing solutions package crop recommendation, fertilizer recommendation and yield prediction together. This paper addresses that gap.

III. MATERIALS AND METHODS

3.1 System overview

The proposed system follows the pipeline shown in the project architecture: data acquisition, data pre-processing, a train/test split, model training, prediction and a GUI application. A user submits soil and weather values through a web form. The crop recommendation model first predicts a suitable crop. That crop is then passed to the fertilizer model and the yield model, so a single submission produces a crop, a fertilizer with a confidence score, and a yield estimate with a confidence score.

3.2 Datasets

Three CSV datasets, obtained from Kaggle and the UCI Machine Learning Repository, support the three tasks (Table

Table 1. Datasets used.

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 2, November 2026
Impact Factor: 8.2

Dataset	Input attributes	Target
Crop recommendation	N, P, K, temperature (°C), humidity (%), pH, rainfall (mm)	Recommended crop (class label)
Fertilizer prediction	Temperature, humidity, soil moisture, soil type (sandy, loamy, black, red, clayey), crop type, nitrogen, potassium, phosphorus	Fertilizer name
Crop yield in Indian states	Crop, crop year, season, state, area (ha), annual rainfall (mm), fertilizer (kg), pesticide (kg), production (t)	Yield (production per unit area)

3.3 Pre-processing

Pre-processing has three steps. First, missing values are replaced and noisy or redundant records are removed (Weka's *ReplaceMissingValues* filter). Second, categorical attributes such as crop type, soil type, season and state are encoded numerically (string attributes are converted to nominal values). Third, numeric attributes are rescaled to the range 0 to 1 with min-max normalisation (Weka's *Normalize* filter); the fertilizer pipeline additionally applies standardisation.

3.4 Feature selection

To remove irrelevant and redundant attributes, the crop recommendation pipeline applies Correlation-based Feature Selection (CFS subset evaluation) with a Best-First search. CFS retains attributes that correlate strongly with the class and weakly with each other, which reduces the number of inputs the farmer has to supply and lowers the risk of over-fitting.

3.5 Prediction models

Crop recommendation. A multilayer perceptron (ANN) is trained on the selected, normalised features to classify the most suitable crop. [State hidden-layer configuration, learning rate and training epochs; the Weka defaults are used if not changed.]

Fertilizer recommendation. A Random Forest classifier predicts the fertilizer name from soil nutrients, climate, soil type and crop type. The final model uses 200 trees, unrestricted depth, three randomly selected features per split, 100% bag size and out-of-bag error estimation; the evaluation model uses 300 trees. The predicted class probability is reported to the user as a confidence score.

Crop yield prediction. An ANN predicts yield from crop, season, state, area, rainfall, fertilizer and pesticide use. Linear Regression serves as a baseline, and a recurrent neural network (RNN) trained with back-propagation through time is also specified in the project design (80% training, 20% testing, Min-Max scaling).

3.6 Evaluation protocol

Classification models are evaluated with cross-validation using accuracy, precision, recall and F1-score:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN);$$

$$\text{Precision} = TP / (TP + FP);$$

$$\text{Recall} = TP / (TP + FN);$$

$$F1 = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}).$$

The fertilizer Random Forest is evaluated with 10-fold cross-validation (seed 1) on the combined data, reporting weighted precision, recall and F-measure. For regression-style outputs the design also specifies MAE, RMSE and the correlation coefficient. [State the number of folds used for the ANN models and the evaluation of the yield model.]

IV. IMPLEMENTATION

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 2, November 2026
Impact Factor: 8.2

The application is a Java web system. The machine learning components use the Weka library, the user interface uses JSP, Servlets, HTML and CSS served by Apache Tomcat, and MySQL (accessed through JDBC) stores datasets, user accounts and prediction results. JFreeChart is used for graphical output and Apache Commons Math for numerical operations. Development was done in Eclipse on Windows 10/11, and the minimum hardware requirement is modest (Intel i3, 2 GB RAM). The interface provides user registration and login, dataset upload for the three datasets, a classifier module, and a crop recommendation page where the user enters N, P, K, temperature, humidity, rainfall and pH and receives the crop, fertilizer and yield outputs.

V. RESULTS AND DISCUSSION

5.1 Model performance

Table 2 summarises the cross-validation results reported for the three tasks.

Table 2. Performance of the proposed models.

Task	Best model	Reported accuracy
Crop yield prediction	ANN	94.70%
Fertilizer recommendation	Random Forest	93.94%
Crop recommendation	[ANN / RNN, to be confirmed]	[insert]

The ANN gave the best yield results, which is consistent with the ability of neural networks to model non-linear interactions between rainfall, inputs and crop outcomes. Random Forest suited the fertilizer task, where the data mixes numeric nutrient values with categorical soil and crop types, and averaging many trees reduces variance. Both results exceed the 90% accuracy target set in the project's acceptance test.

5.2 Functional testing

The system was tested with ten cases covering dataset loading, pre-processing, feature selection, crop recommendation, fertilizer prediction, yield prediction, model accuracy, full workflow integration, interface usability and performance on a large dataset. All ten cases passed, and no errors were observed in the end-to-end workflow.

5.3 Practical value

By returning a crop, a fertilizer and a yield estimate together, the system reduces the effort a farmer needs to interpret data and supports more efficient fertilizer use, which can lower cost and environmental impact.

VI. LIMITATIONS

Prediction quality depends on the size, quality and age of the public datasets, and the datasets may not reflect current local conditions.

The models are trained on tabular data and do not use real-time sensor feeds, satellite imagery, pests or disease information.

The yield value is derived from region-level records, so it may not transfer to an individual farm.

Neural networks are less interpretable than tree-based models, which matters for farmer trust.

Reported accuracy comes from cross-validation on public data; field validation with real farms has not yet been carried out.

VII. CONCLUSION AND FUTURE WORK

This paper presented an integrated machine learning framework and web application for crop recommendation, fertilizer recommendation and crop yield prediction. Combining feature selection with ANN and Random Forest models, the system reached 94.70% accuracy for yield prediction and 93.94% for fertilizer recommendation, and it

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 2, November 2026 Impact Factor: 8.2

delivers the three outputs through one farmer-friendly interface. Future work includes real-time IoT sensor integration, weather forecast feeds, additional crops and regions, regional-language support, explainable AI for the recommendations, and field trials with farmers.

REFERENCES

[1] T. Vinod, N. C. Reddy, R. S. Reddy and A. A. Reddy, "Crop yield prediction using machine learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 12, no. 4, Apr. 2023.
[2] S. Shariff, S. R. B., R. O. G., P. H. and P. K., "Crop recommendation using machine learning techniques," *IJERT*, ICEI-2022 Conference Proceedings, 2022.
[3] P. R. Hegde and A. R. Ashok Kumar, "Crop yield and price prediction system for agriculture application," *IJERT*, vol. 11, no. 7, Jul. 2022.
[4] N. Bali and A. Singla, "Emerging trends in machine learning to predict crop yield and study its influential factors: A survey," *Archives of Computational Methods in Engineering*, pp. 1–18, 2022.
[5] V. K. Gunjan et al., "Prediction of agriculture yields using machine learning algorithms," in *Proc. 2nd Int. Conf. Recent Trends in Machine Learning, IoT, Smart Cities and Applications (ICMISC 2021)*, Springer Singapore, 2022.
[6] D. J. Reddy and M. R. Kumar, "Crop yield prediction using machine learning algorithm," in *Proc. 5th Int. Conf. Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2021.
[7] S. Agarwal and S. Tarar, "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms," *Journal of Physics: Conference Series*, vol. 1714, no. 1, IOP Publishing, 2021.
[8] A. Morales and F. J. Villalobos, "Using machine learning for crop yield prediction in the past or the future," *Frontiers in Plant Science*, vol. 14, p. 1128388, 2023.
[9] S. V. Joshua et al., "Crop yield prediction using machine learning approaches on a wide spectrum," *Computers, Materials & Continua*, vol. 72, no. 3, 2022.
[10] M. K. Senapaty, A. Ray and N. Padhy, "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture," *Computers*, vol. 12, no. 3, p. 61, 2023.
[11] 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, 2024.
[12] T. S. T. Tanaka et al., "Can machine learning models provide accurate fertilizer recommendations?," *Precision Agriculture*, vol. 25, no. 4, pp. 1839–1856, 2024.
[13] S. Okwii, "NPK soil fertility measurement system and crop recommendation based on machine learning," 2023.
[14] S. Sundaresan et al., "Machine learning and IoT-based smart farming for enhancing the crop yield," *Journal of Physics: Conference Series*, vol. 2466, IOP Publishing, 2023.
[15] L. Gottemukkala et al., "Sustainable crop recommendation system using soil NPK sensor," *E3S Web of Conferences*, vol. 430, EDP Sciences, 2023.