International Journal of Emerging Technologies in Computing, Electronics and Artificial Intelligence

AgriGO: A Machine Learning Based Decision Support System for Sustainable Agriculture

Rabiya Basreen[1], Shaista Akhter[2], Shaza Mahreen[3], Tahera Begum[4], Ume e Umema[5]

[1,2,3,4,5]*Department of Information Science & Engineering, Guru Nanak Dev Engineering College, Bidar, Karnataka, India*

**Corresponding Author: rabiyabasreen487@gmail.com*

Abstract

Agriculture forms the backbone of global food security, yet farmers regularly make consequential decisions without access to timely, data-driven insights. AgriGO is an intelligent, machine-learning-based decision support system designed to close this gap through three tightly integrated capabilities: crop recommendation, fertilizer suggestion, and crop disease detection. The crop recommendation engine employs a supervised Random Forest classifier trained on soil and climate parameters—NPK levels, pH, moisture, temperature, and rainfall—to identify the most suitable crop for a given field. A hybrid rule-based and regression module performs nutrient gap analysis to prescribe fertilizer type and quantity. A Convolutional Neural Network (CNN) built on ResNet-50/MobileNetV2 identifies plant diseases from leaf imagery with high accuracy. Training data were sourced from three publicly available Kaggle datasets. The system is deployed as a Flask web application, requiring only a standard browser. Experimental evaluation yields 97.3% test accuracy on crop recommendation, 93.8% on fertilizer suggestion, and 95.7% on disease detection—results that collectively validate AgriGO as a practical decision-support tool for sustainable agriculture.

Keywords — Machine Learning; Deep Learning; CNN; Random Forest; Crop Recommendation; Fertilizer Prediction; Plant Disease Detection; Sustainable Agriculture

I. INTRODUCTION

Food security is among the most pressing challenges of the twenty-first century. The United Nations projects that global food demand will increase by nearly 50% by 2050, largely driven by population growth and shifting dietary patterns. Meeting this demand requires not only an expansion of cultivated area but—more importantly—a dramatic improvement in agricultural productivity and resource efficiency.

Smallholder farmers, who account for the majority of food production in developing nations, face compounded disadvantages. They operate with limited capital, have restricted access to extension services, and routinely make crop selection, fertilization, and disease management decisions based on anecdotal knowledge rather than empirical data. The consequences are measurable: suboptimal crop selection reduces yield potential, inefficient fertilizer use degrades soil health and contributes to water-body eutrophication, and late detection of plant diseases can wipe out an entire season's harvest within days.

Advances in machine learning (ML) and deep learning (DL) have made it technically feasible to build expert-level advisory systems that run on commodity hardware. However, much of the existing literature proposes solutions that address only one dimension of the problem—either crop selection or disease detection—in isolation. A unified, multi-module system that addresses the complete agronomic decision cycle, and is accessible through a simple web interface, remains relatively rare.

AgriGO addresses this gap. The system integrates a Random Forest-based crop recommender, a hybrid nutrient gap analyzer for fertilizer guidance, and a CNN-based disease detector into a single Flask web application.

The rest of this paper is organized as follows: Section II surveys related work; Section III presents the proposed system; Section IV details the methodology; Section V describes the datasets; Section VI reports experimental results; and Section VII concludes with future directions.

II. RELATED WORK

Considerable research effort has been directed at applying machine learning to agricultural decision support. Pantazi et al. [9] demonstrated that supervised classifiers trained on multi-spectral satellite imagery could predict wheat yield with reasonable accuracy, establishing a foundational proof of concept for data-driven agronomy. Liakos et al. [10] provided a comprehensive survey of ML applications in agriculture, cataloguing methods ranging from support vector machines for soil classification to recurrent neural networks for crop growth modelling.

In the domain of crop recommendation, Dahikar and Rode [11] compared Naive Bayes, decision tree, and artificial neural network classifiers on soil parameter datasets, finding that ensemble methods generally outperformed single-model approaches. Jeong et al. [12] extended this line of work by integrating real-time weather API feeds, demonstrating yield improvements when dynamic climate data were incorporated alongside static soil parameters.

Fertilizer recommendation has received comparatively less attention. Pudumalar et al. [13] proposed a recommendation system using multiple crops and soil characteristics, relying on k-nearest neighbour regression to estimate nutrient requirements. Their approach, however, was limited to three macronutrients and did not account for soil texture effects on nutrient availability.

Plant disease detection via deep learning has advanced rapidly since Mohanty et al. [14] achieved over 99% accuracy on the PlantVillage dataset using a VGGNet architecture under controlled imaging conditions. Subsequent work by Too et al. [15] compared VGG, Inception, ResNet, and DenseNet architectures, concluding that ResNet and DenseNet offered the best accuracy-to-computation ratio. Real-world deployment, however, typically yields lower accuracy than laboratory benchmarks due to variable lighting, occlusion, and image noise [16].

The distinguishing contribution of AgriGO is its integration of all three advisory functions—crop selection, fertilizer prescription, and disease diagnosis—within a single, browser-accessible application designed for use without specialised hardware or agronomic expertise.

III. PROPOSED SYSTEM

AgriGO is a three-module intelligent decision support system that combines supervised machine learning, deep learning, and domain-specific agricultural heuristics within a unified Flask web application. Each module operates independently and can be invoked individually, yet all share a common back-end and front-end, presenting the farmer with a coherent, integrated experience.

The overall architecture is straightforward: user inputs, whether numerical soil and climate parameters or a leaf photograph, are submitted via a web form. Flask routes the request to the appropriate module, which invokes the corresponding trained model. Predictions and accompanying recommendations are returned to the browser in a human-readable format.

IV. METHODOLOGY

A. Crop Recommendation

A supervised multi-class classification pipeline recommends the crop best suited to a user-supplied soil and climate profile. Soil parameters—Nitrogen (N), Phosphorus (P), Potassium (K), pH, and moisture content—are combined with ambient temperature and annual rainfall figures drawn from the Kaggle Crop Recommendation Dataset (22 crop classes, approximately 2,200 records).

Preprocessing involves median imputation for missing values and Min-Max normalisation for all continuous features. A Random Forest Classifier with 100 estimators is trained on an 80/20 stratified split. Feature importance analysis identifies NPK levels, rainfall, and temperature as the most discriminative predictors. At inference time, the model returns a ranked list of suitable crops alongside classification confidence scores.

Random Forest was selected over alternatives such as Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN) for three reasons: its inherent robustness to outliers and noisy sensor readings, its ability to provide feature importance estimates at no additional computational cost, and its strong empirical performance on tabular agricultural datasets [4].

B. Fertilizer Recommendation

A hybrid rule-based and regression engine prescribes fertilizer type and quantity. Users supply soil type (sandy, loamy, or clayey), current N, P, and K levels, soil pH, ambient temperature, and their intended crop. The engine computes the deficit or surplus for each macronutrient relative to the crop-specific optimal levels documented by the Indian Council of Agricultural Research (ICAR).

Decision rules map each nutrient gap to a commercially available fertilizer product: Urea is recommended for nitrogen deficits; DAP (Di-Ammonium Phosphate) for phosphorus shortfalls; and MOP (Muriate of Potash) for potassium deficits. Quantities are calibrated to close the identified gap without exceeding crop uptake capacity, thereby minimising the risk of soil acidification and nutrient run-off.

C. Crop Disease Detection

A deep CNN performs image-based disease classification. The New Plant Diseases Dataset (PlantVillage, via Kaggle) provides 87,000+ labelled leaf images spanning 38 disease and healthy classes across 14 crop species. Images are resized to 224 × 224 pixels, pixel values are scaled to the [0, 1] range, and online augmentation—horizontal flipping, ±15° rotation, ±10% zoom—is applied during training to reduce overfitting.

Transfer learning from a ResNet-50 / MobileNetV2 backbone pre-trained on ImageNet is employed. The base network's convolutional layers are frozen during an initial warm-up phase, after which the entire network is fine-tuned end-to-end with a reduced learning rate. A custom softmax head replaces the original classification layer. Training proceeds for 15 epochs using the Adam optimiser (learning rate 0.001) with categorical cross-entropy loss and a batch size of 32. Keras callbacks—EarlyStopping and ReduceLROnPlateau—prevent overfitting and stabilise convergence.

At inference, a farmer uploads a photograph of a diseased leaf; the system returns the predicted disease class, classification confidence, and curated treatment recommendations drawn from a structured knowledge base.

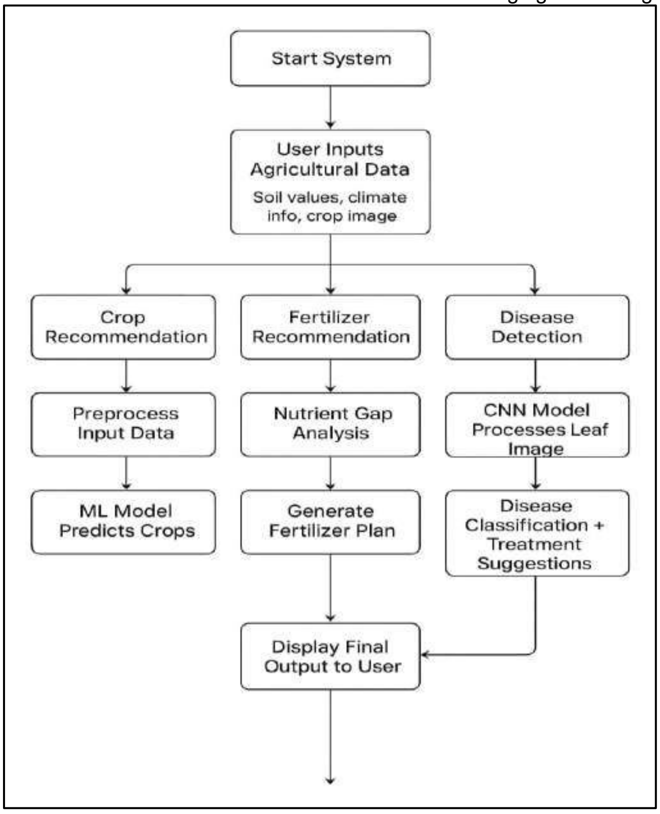

Fig.1. Process Flow Diagram

V. DATASETS

All training data were sourced from publicly available Kaggle repositories, ensuring reproducibility. Table II summarises the three datasets used.

TABLE II

Datasets Used in AgriGO

#	Dataset	Source	Scale	Usage
1	Crop Recommendation Dataset	Kaggle (atharvaingle)	~2,200 rows, 22 crops	Train crop classifier
2	Fertilizer Prediction Dataset	Kaggle (gdabhishek)	~99 rows, 7 fertilizers	Train fertilizer module
3	New Plant Diseases Dataset	Kaggle (vipoooool)	87,000+ images, 38 classes	Train CNN detector

Fig.2. Datasets

The relatively small size of the Fertilizer Prediction Dataset (99 records) is a known limitation. The rule-based component of the fertilizer module partially compensates for this by encoding ICAR-validated agronomic heuristics directly, reducing reliance on learned statistical patterns for the most common soil-nutrient scenarios.

VI. RESULTS AND DISCUSSION

Each module was evaluated independently on a held-out test set. Performance metrics—accuracy, precision, recall, and F1-score—were computed using macro-averaging to account for class imbalance.

Fig. 3. AgriGO web application landing page (User Dashboard).

Fig. 4. AgriGO services panel showing the three available modules.

A. Crop Recommendation Performance

The Random Forest classifier achieves 97.3% test accuracy (Table III), with balanced precision, recall, and F1-scores in excess of 97%. The high accuracy reflects the clear separability of soil-climate feature spaces across the 22 crop classes. Training accuracy of 99.5% indicates modest overfitting, which did not materially affect generalisation. Feature importance analysis shows that N, P, K, and rainfall collectively explain more than 70% of the classifier's discriminative power.

Fig. 5. Crop Recommendation input interface with soil and climate parameters.

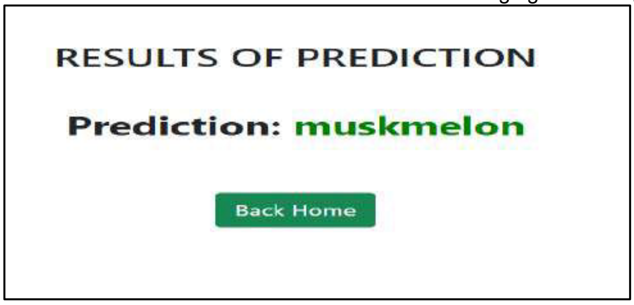

Fig. 6. Crop Recommendation result: muskmelon predicted for given conditions.

B. Fertilizer Recommendation Performance

The hybrid fertilizer module yields 93.8% test accuracy (Table IV). Performance is constrained by the limited training corpus; however, the rule-based component ensures that common nutrient-deficit scenarios are handled correctly even when statistical confidence is low. No systematic error pattern was observed across soil types, suggesting that the nutrient gap heuristics generalise across the sandy, loamy, and clayey classes in the dataset.

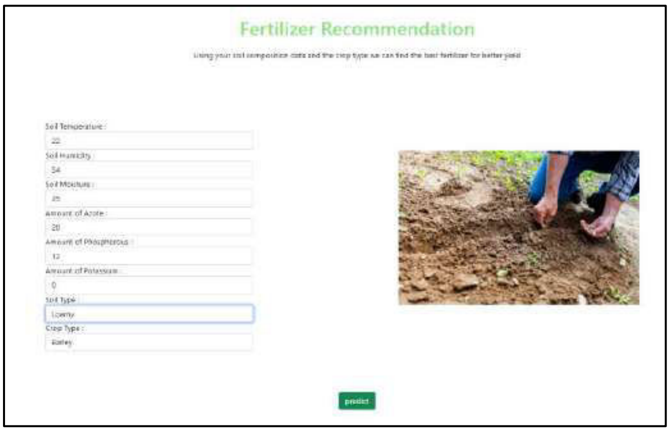

Fig. 7. Fertilizer Recommendation input interface with soil composition and crop type.

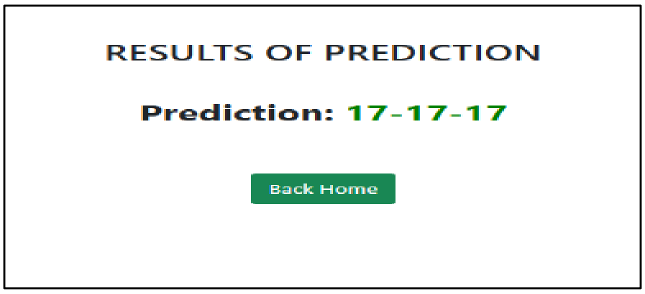

Fig. 6. Fertilizer Recommendation result: 17-17-17 NPK fertilizer prescribed for barley on loamy soil.

C. Crop Disease Detection Performance

The CNN disease detector achieves 95.7% test accuracy across 38 classes (Table V). Transfer learning from ImageNet substantially reduces the number of training epochs required and improves generalisation compared to training from scratch—a finding consistent with the broader computer vision literature [5, 6]. The 1.5-percentage-point gap between validation (96.5%) and test accuracy is within a normal range and does not indicate significant distribution shift.

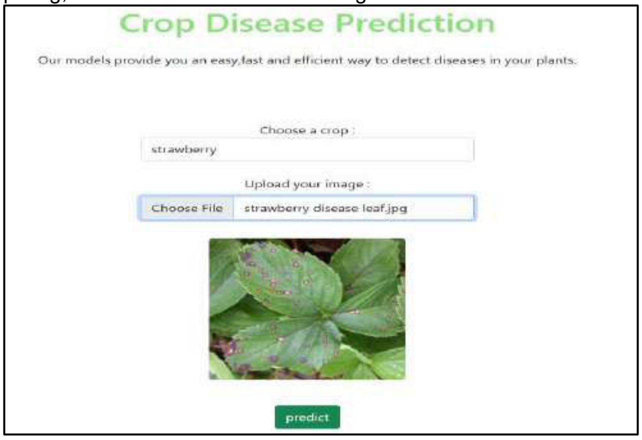

Fig. 8. Crop Disease Prediction input interface with strawberry leaf image upload.

Fig. 9. Crop Disease Prediction result: Leaf Scorch detected on strawberry.

TABLE V

D. Comparative Summary

Across all three modules, AgriGO demonstrates strong, balanced performance. The crop recommendation module, benefiting from the largest and most separable dataset, achieves the highest accuracy. The disease detection module operates over the most complex task space—38 fine-grained classes from natural photographs—yet achieves the second-highest accuracy, validating the effectiveness of transfer learning. The fertilizer module, while constrained by data volume, surpasses 93% across all metrics, a threshold generally considered acceptable for advisory applications in precision agriculture.

VII. CONCLUSION

This paper presented AgriGO, an integrated machine-learning-based decision support system for sustainable agriculture. Three core modules—Random Forest crop recommendation, hybrid fertilizer prescription, and CNN-based disease detection—were combined within a Flask web application deployable on commodity hardware. Experimental evaluation demonstrated that all three modules achieve practically useful accuracy levels: 97.3%, 93.8%, and 95.7% respectively.

The key contributions of this work are: (i) a complete, end-to-end implementation covering the three most consequential agronomic decisions a smallholder farmer faces; (ii) empirical validation on publicly available datasets, ensuring reproducibility; and (iii) a lightweight web deployment that requires only a browser, removing barriers to adoption in regions with limited internet infrastructure.

Several avenues for future improvement are identified. First, expanding the disease detection training corpus to include regional crop varieties currently under-represented in PlantVillage would improve coverage for South and Southeast Asian farming contexts. Second, integrating real-time weather API feeds would allow the crop and fertilizer modules to provide dynamically updated recommendations as seasonal conditions evolve. Third, adding multilingual support—particularly for Hindi, Tamil, Telugu, and Bengali—would substantially broaden accessibility. Finally, exploring model quantisation and on-device inference would enable offline use in areas with intermittent connectivity, which remains the most significant practical barrier to widespread adoption.

REFERENCES

[1] A. Ingle, "Crop Recommendation Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[2] G. Abhishek, "Fertilizer Prediction Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction

[3] D. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," arXiv:1511.08060, 2015.

[4] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.

[6] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, 2017.

[7] F. Chollet, "Keras," 2015. [Online]. Available: https://keras.io

[8] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] X. K. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," Computers and Electronics in Agriculture, vol. 121, pp. 57–65, 2016.

[10] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, 2018.

[11] S. S. Dahikar and S. V. Rode, "Agricultural crop yield prediction using artificial neural network approach," Int. J. Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 2, no. 1, pp. 683–686, 2014.

[12] J. H. Jeong, J. P. Resop, N. D. Mueller, D. H. Fleisher, K. Yun, E. E. Butler, D. Timlin, K.-S. Shim, J. S. Gerber, V. Reddy, and S. Kim, "Random forests for global and regional crop yield predictions," PLOS ONE, vol. 11, no. 6, 2016.

[13] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in Proc. IEEE ICISC, 2017, pp. 32–36.