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, August 2026
Impact Factor: 8.2

ML Framework for Profitable Crop Yield Based on Climatic and Economic Factors

Kingshuk Ghosh[1], Sahil Singh[2], Dr. Usha G[3]
Dept. of Computing Technologies[1-3]
SRM Institute of Science and Technology, Tamil Nadu, India

Abstract: *Agricultural planning is inherently uncertain: soil composition varies across seasons, and market prices for produce are rarely predictable at the time planting decisions are made. Farmers routinely face this uncertainty with limited analytical support, which motivates the development of the Agricultural Decision Support System (ADSS) presented in this work. The ADSS assists farmers by combining agronomic and economic data to generate crop recommendations, evaluating nitrogen, phosphorus, and potassium content, temperature, humidity, rainfall, and soil pH to identify crops suited to given field conditions and likely to be commercially viable. Crop recommendations are generated using an XGBoost classifier, achieving 98.86% accuracy across 22 crop categories. In addition to crop suitability, the system estimates expected profitability for the recommended crop; factors such as adverse weather and pest incidence remain outside the scope of the model. A deep learning autoencoder is incorporated to denoise the input feature space while preserving informative structure, improving downstream model efficiency. The resulting system operates with sub-millisecond inference latency, making it suitable for near-real-time deployment. Recommendations are accompanied by natural-language explanations, allowing users to understand the reasoning behind each suggestion. Overall, the ADSS functions as a data-driven advisory tool that jointly considers agronomic suitability and economic outcome, positioning it as a practical component of precision agriculture powered by machine learning and artificial intelligence.*

Keywords: Precision Agriculture, Machine Learning, Crop Recommendation, XGBoost, Agricultural Decision Support System, Artificial Intelligence.

I. INTRODUCTION

Feeding a growing global population under increasingly variable environmental conditions remains one of the central challenges facing modern agriculture. While farming continues to sustain billions of people, rising demand combined with climate variability, soil degradation, unpredictable rainfall, and limited access to technology has exposed the limits of conventional practice. Mismatches between crop choice and field conditions reduce yields and impose measurable economic losses on producers. Crop selection has traditionally relied on accumulated experience and inherited knowledge. Such knowledge remains valuable, but soil fertility is not static; it varies from season to season and from field to field. Even an experienced farmer cannot easily synthesize dozens of interacting variables, including pH, potassium content, and humidity, without analytical support. The resulting mismatch between crop and field conditions contributes to wasted resources and reduced income.

Advances in information technology and artificial intelligence have introduced new possibilities for addressing this gap. Precision agriculture, which relies on data rather than intuition alone, has gained increasing traction [7], [9]. Decision Trees, Random Forests, and Gradient Boosting methods have been evaluated for crop recommendation, with several systems reporting accuracy in the high nineties [1], [10], a level of precision not attainable through manual estimation alone. However, many existing systems evaluate only agronomic suitability without incorporating economic outcome. This is a significant limitation, since farming operates as an economic activity in which profitability, not merely biological viability, determines the value of a recommendation. Recommending a crop that grows well but yields low market returns or requires high input costs provides limited practical benefit. Recent systems have begun to incorporate profit prediction alongside environmental suitability [10], extending the scope of recommendation from "will this crop grow" to "will this crop be profitable." A further consideration is user trust. Recommendation systems that provide only a categorical output, without justification, offer limited practical utility if the reasoning remains opaque to the end user. Incorporating explanation — for example, "rice is recommended because the soil exhibits high nitrogen content and favorable humidity, with phosphorus supplementation advised" — closes this gap by aligning the system's output with the reasoning process of a human advisor. This direction is consistent with recent work in explainable artificial intelligence [2].

This work presents an Agricultural Decision Support System (ADSS) that integrates these considerations. The system accepts environmental inputs — nitrogen, phosphorus, potassium, pH, temperature, humidity, and rainfall — and produces both a crop recommendation and a profit estimate. Crop recommendation is performed using an XGBoost classifier, while profit estimation is performed using an XGBoost regression model. A deep learning autoencoder is used for feature engineering to reduce noise in the input space and improve downstream efficiency. A distinguishing component of the proposed system is an AI advisory module: an autoregressive language model is deployed locally to generate natural-language explanations, fertilizer recommendations, irrigation guidance, and management strategies, without transmitting farm data to external cloud services. This design choice preserves data locality while approximating the experience of consulting a knowledgeable agronomist. The principal contributions of this work are as follows: (i) a machine-learning-based system for crop recommendation; (ii) profit prediction to ground recommendations in economic reality; (iii) autoencoder-based feature extraction to improve model efficiency; (iv) an autoregressive advisory module for generating explanations and management guidance; and (v) evaluation using standard agricultural datasets [8]. Experimental results indicate 98–99% accuracy for crop recommendation and strong regression performance for profit estimation, suggesting that a data-driven approach can meaningfully improve agricultural decision-making and resource allocation. The remainder of this paper is organized as follows. Section II reviews prior work on machine learning applications in agricultural systems. Section III describes the architecture of the proposed decision support system. Section IV details the experimental procedure and dataset used in this study. Section V presents the experimental results and evaluates system performance. Section VI concludes the paper and outlines directions for future research.

II. LITERATURE REVIEW

Considerable prior work has explored the use of machine learning and artificial intelligence to characterize environmental conditions, forecast crop performance, and support cultivation decisions. This section reviews relevant literature across several themes: machine learning for crop recommendation, precision agriculture and the Internet of Things (IoT), economic analysis in crop selection, decision support systems, explainability, and remaining research gaps.

A. Machine Learning-Based Crop Recommendation Systems

Machine learning has become a widely adopted approach for crop recommendation, with the central objective of identifying relationships between soil nutrients, climate variables, and crop outcomes. This typically involves providing a model with inputs such as nitrogen, phosphorus, potassium, rainfall, temperature, and humidity, and training it to recommend a suitable crop.

One study constructed a recommendation system using these inputs and found that the machine learning approach substantially outperformed conventional selection methods [1], consistent with the expectation that well-tuned models can identify patterns across multiple simultaneous variables that manual assessment may overlook.

Another study proposed a system employing supervised learning techniques, including Support Vector Machines and Random Forest classifiers [10], reporting that ensemble approaches perform well on agricultural data and yield accurate predictions within the reported experimental setting.

The underlying relationship can be expressed formally as:

$$\text{Crop} = f(x_1, x_2, x_3, \ldots, x_n)$$

where each x_i represents an input parameter such as temperature, humidity, rainfall, soil nutrients, or land area, and the output is the predicted crop label. Ensemble learning algorithms such as Gradient Boosting and Random Forest are commonly used in modern systems, owing to their effectiveness in capturing the nonlinear relationships that govern crop growth.

B. Precision Agriculture

A related line of research integrates machine learning with real-time sensor data as part of precision agriculture, in which field-deployed sensors continuously measure variables such as soil moisture, temperature, and humidity to inform cultivation decisions.

One study proposed an IoT-based system that collects environmental data and applies machine learning models to assist crop selection [10]. A related study developed a smart farming system using IoT sensors combined with similar models to enable rapid response to changing weather conditions [9].

Real-time feedback of this kind offers a meaningful advantage over reliance on historical averages, providing field-specific information rather than generalized estimates. The practical benefit depends on sensor reliability and the availability of supporting infrastructure, though the overall trajectory favors continuous, sensor-informed decision-making. A further study described an IoT-based monitoring system that uses sensor data to assess crop suitability [4]. Collectively, this body of work points toward data-driven cultivation becoming standard practice rather than the exception.

C. Economic Analysis in Crop Selection

A recurring limitation in early crop recommendation systems is their exclusive focus on environmental suitability, without accounting for economic considerations such as input cost, market price, and expected return. A crop that performs well agronomically but yields limited market value, or requires costly inputs, may not represent a sound recommendation.

One study addressed this limitation by developing a machine learning model that used soil and environmental factors to estimate crop yield [5], providing farmers with a yield projection that supports planning and approximate income estimation.

Regression is a common approach for modeling yield:

$$\text{Yield} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where the x_i are environmental parameters. Model performance is commonly evaluated using Mean Squared Error (3). Yield prediction alone, however, addresses only part of the decision problem; profitability remains the primary concern for farmers. In agricultural economics, Return on Investment (ROI) is a standard measure of profitability:

$$\text{ROI} = [(\text{Revenue} - \text{Cost\_total}) / \text{Cost\_total}] \times 100$$

Incorporating this calculation into a recommendation system extends its function beyond agronomic suitability toward genuine economic decision support.

D. Decision Support Systems in Agriculture

Decision support systems in agriculture integrate these analytical components into a unified framework, assisting farmers with crop selection, irrigation planning, and resource allocation.

One example is a system that evaluates soil fertility and environmental conditions to recommend suitable crops [8], applying machine learning to agricultural data to generate productivity-oriented suggestions. A further study demonstrated that data-driven decision-making extends to a broader range of applications, showing that machine learning models can assist farmers in interpreting environmental data to select appropriate crops [10].

These systems are most effective when they reduce the uncertainty inherent to agricultural planning. A well-designed decision support system can provide automated insights that reduce planning time and help avoid costly errors.

E. Research Gap

As previously noted, decision support systems function as integrative frameworks that combine analytical components to assist with crop selection, irrigation, and related planning tasks. Most existing systems, however, address environmental suitability and economic outcome as separate concerns rather than a unified recommendation, and rarely provide an explanation that a non-technical end user can act on. This gap — the absence of a system that jointly models agronomic fit, expected profitability, and interpretable guidance — motivates the ADSS proposed in this work.

III. SYSTEM ARCHITECTURE

The Agricultural Decision Support System (ADSS) is designed to extend beyond simple crop identification by jointly addressing agronomic suitability and economic outcome. The system integrates several components: machine learning models for prediction, feature engineering for data refinement, and an advisory module that generates natural-language explanations for its outputs.

The architecture consists of a sequence of modules through which information is processed: data acquisition, preprocessing, feature engineering, crop prediction, cost estimation, return-on-investment analysis, and a final AI-based advisory component that interprets the results for the end user. Inputs include soil nutrients and climate conditions, as well as land area, soil type, and prevailing market prices for candidate crops, enabling the system to generate recommendations that reflect both agronomic feasibility and economic value.

A. Data Acquisition Module

The data acquisition module collects the primary environmental inputs: nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and rainfall. The relevance of these variables to crop suitability is well established in prior work, which informed the selection of these input variables [1], [10].

B. Data Preprocessing Module

Raw agricultural data typically requires preprocessing prior to model training. The preprocessing module addresses missing values, scales numerical features to prevent any single variable from dominating model training, and encodes categorical variables such as soil type and crop label into a numerical form. Normalization is particularly important: rainfall values, which may range from a few millimeters to several hundred, would otherwise overshadow variables such as pH, which vary within a narrower range. The dataset is partitioned into training and testing subsets to provide an unbiased assessment of model performance, following preprocessing conventions established in comparable agricultural prediction systems [9].

C. Feature Engineering Module

The feature engineering module is designed to focus model attention on the most informative variables. A deep learning autoencoder is used for this purpose, compressing the input feature space into a smaller representation that retains essential information while discarding noise and redundancy. This approach captures the underlying structure of the data without being unduly influenced by minor fluctuations. The use of deep learning for this purpose reflects a broader trend in agricultural systems, motivated by the complexity of environmental datasets, which often contain patterns not easily captured by simpler methods [7].

D. Crop Prediction Module

The crop prediction module generates the primary recommendation using an XGBoost classifier, a gradient boosting algorithm well suited to structured data such as soil and climate measurements. The model uses soil nutrients,

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, August 2026
Impact Factor: 8.2

temperature, rainfall, humidity, and land area to determine the most suitable crop. XGBoost is widely adopted for this task due to its ability to model nonlinear relationships effectively and scale to larger datasets [9]. The module outputs a ranked list of candidate crops with associated confidence scores, reflecting that more than one viable option may exist for a given set of conditions.

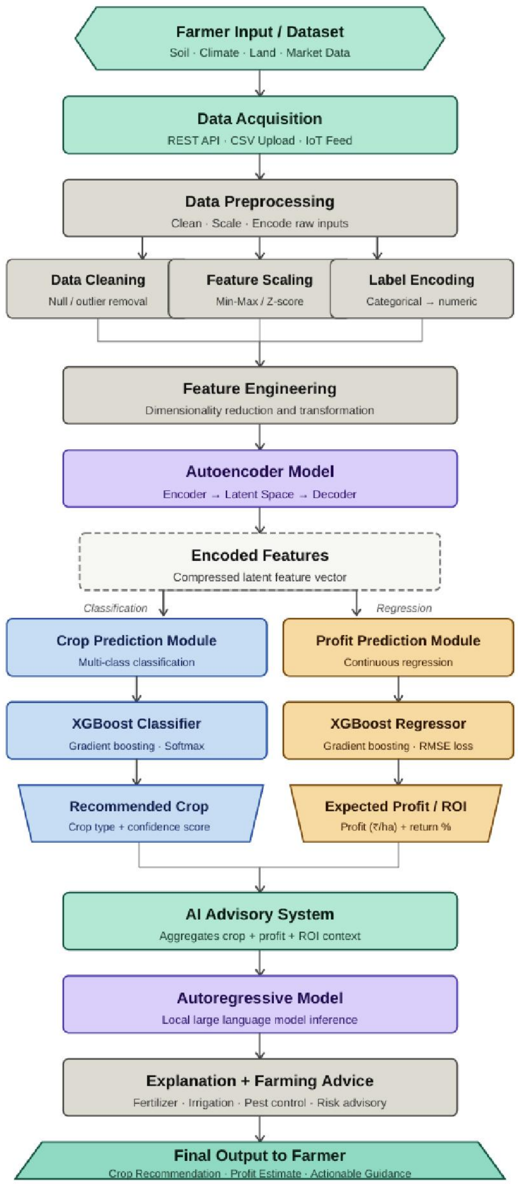

Fig. 1. System Architecture of ADSS

E. Cost Estimation Module

A crop recommendation is more actionable when accompanied by an estimate of cultivation cost. The cost estimation module calculates expected expenses based on soil type, land area, and estimated requirements for fertilizer, irrigation, labor, and land preparation. This estimation is necessarily approximate, as real-world costs vary considerably due to regional labor rates, input pricing, and cultivation practices. The objective is to provide a reasonable estimate grounded in soil characteristics and land size rather than a generic figure. Economic analysis is often absent from crop recommendation systems despite its relevance, given that farming constitutes an economic activity for the end user [9].

F. Revenue and ROI Analysis Module

Following cost estimation, the system calculates expected revenue and return on investment. Expected revenue is derived by multiplying projected yield by current market price, introducing an additional source of uncertainty due to market price volatility. The ROI formula, which relates revenue to cost, produces a percentage indicating expected profitability. While simple, this metric ensures that economic considerations are incorporated directly into the recommendation process rather than treated as secondary. Including ROI shifts the system's focus toward profitability rather than yield alone.

G. AI Advisory Module

The AI advisory module is central to the system's usability, as it converts model output into an interpretable explanation rather than a raw prediction. This module uses a locally deployed large language model, executed through Ollama, to generate explanations and practical guidance. Recommendations are accompanied by an explanation referencing the specific environmental factors that informed the decision, along with suggested fertilizers, irrigation schedules, or management practices. High model accuracy alone does not guarantee adoption; without an understandable rationale, users may be reluctant to act on a recommendation. Explainability has been shown to directly influence user trust and adoption in related work [2].

H. System Workflow

The overall workflow proceeds as follows: data is collected, cleaned, and transformed; the machine learning models generate crop and profit predictions; and the AI advisory module then interprets these outputs and produces actionable guidance. While the pipeline is not highly complex, it addresses three core requirements: environmental compatibility, economic feasibility, and interpretability. Practical effectiveness depends on the successful integration of these components at the point of decision-making, which the system architecture is designed to support.

IV. EXPERIMENTAL SETUP

This section describes the dataset used in this study, the preprocessing and feature engineering procedures applied, the configuration of the two XGBoost models, and the implementation environment.

A. Dataset Description

The dataset used in this study, CropRecommendation.csv, was obtained from Kaggle and is widely used in this research area, allowing results to be compared with prior work. It contains standard environmental variables: nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and rainfall, which are commonly regarded as primary drivers of crop suitability. Comparable datasets have been used in several earlier studies [1], [10].

The dataset includes 22 distinct crop categories, each represented by exactly 100 samples: apple, banana, blackgram, chickpea, coconut, coffee, cotton, grapes, jute, kidneybeans, lentil, maize, mango, mothbeans, mungbean, muskmelon, orange, papaya, pigeonpeas, pomegranate, rice, and watermelon. This balanced distribution reduces the risk of classifier bias toward more frequently represented crops and ensures the model receives adequate exposure to each class during training

DOI: 10.48175/IJARSCT-38126

ISSN: 2581-9429
Volume 6, Issue 2, August 2026
Impact Factor: 8.2

B. Data Exploration

Prior to model training, exploratory analysis was conducted on the dataset. The dataset comprises 2200 rows with no missing values across any column, eliminating the need for imputation. Nutrient values span a wide range: nitrogen from 0 to 140, phosphorus from 5 to 145, and potassium from 5 to 205. Temperature and rainfall similarly cover a broad range. This diversity supports generalization across climates and soil conditions rather than overfitting to a narrow range of values. The balanced crop distribution described above was confirmed during this exploratory phase.

C. Data Preprocessing

Standard preprocessing steps were applied. Environmental features were separated from crop labels, crop names were encoded numerically using a label encoder, and all numerical features were scaled using a standard scaler. Scaling is essential in this context: without it, variables such as rainfall, which range into the hundreds, would dominate variables such as pH, which occupy a narrow range. The dataset was split 80/20, yielding 1760 training samples and 440 testing samples, providing a substantial holdout set for evaluation.

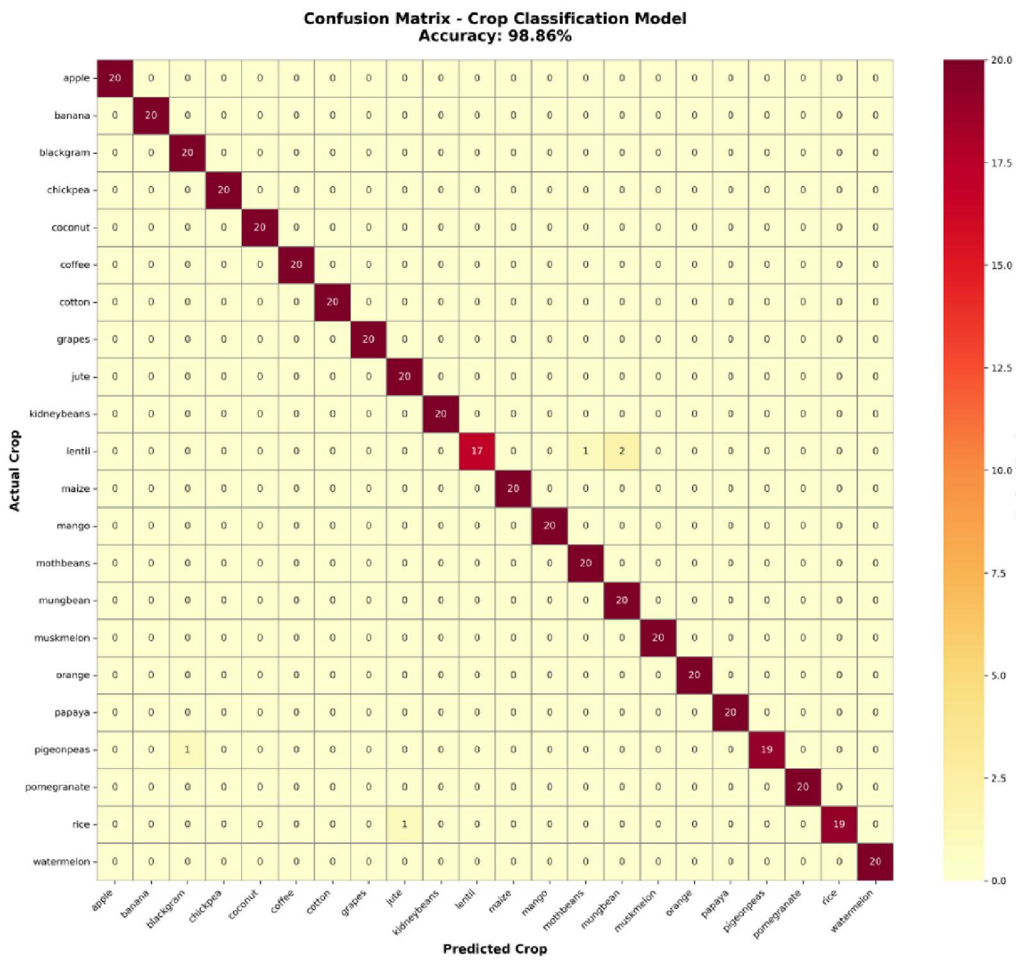

Fig. 2. Prediction performance across all 22 crop categories

DOI: 10.48175/IJARSCT-38126

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, August 2026
Impact Factor: 8.2

D. Feature Engineering using Autoencoder

Rather than providing all seven raw features directly to the classifier, a deep learning autoencoder was used to compress them into four latent features. This step was intended to reduce noise and redundancy so that the classifier could focus on the most informative components of the data. The autoencoder was trained for 100 epochs using several dense layers, learning to reconstruct the original inputs with minimal error.

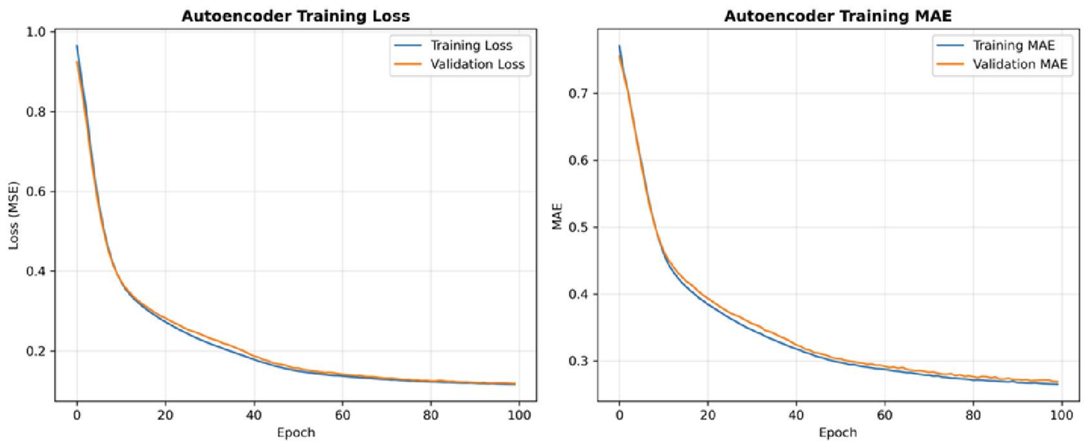

Fig. 3. Autoencoder training history

The final reconstruction loss (mean squared error) was approximately 0.1157 on the training set and 0.1183 on the test set, with a mean absolute error of 0.2689 on the test set. These values indicate that the autoencoder captured the essential structure of the data with limited information loss, consistent with prior findings suggesting that deep feature extraction can improve downstream performance in agricultural datasets [7].

TABLE I: FINAL PERFORMANCE METRICS OF THE AUTOENCODER

Metric	Value
Training Loss (MSE)	0.1157
Testing Loss (MSE)	0.1183
Testing MAE	0.2689

E. Crop Prediction using XGBoost Classifier

Using the compressed feature representation, an XGBoost classifier was trained to generate crop recommendations. XGBoost, a gradient boosting algorithm, is generally effective on structured, tabular data of this kind. Default hyperparameters were largely retained, as the primary objective was to evaluate the overall system architecture rather than perform extensive hyperparameter tuning.

The classifier achieved 98.86% accuracy on the test set, with macro- and weighted-average precision and recall both at 0.99, indicating consistently low error rates across all 22 crop categories. A 5-fold cross-validation yielded a mean accuracy of 97.2% with a standard deviation of 1.3%, indicating stable performance independent of a particular data split.

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, August 2026 Impact Factor: 8.2

Of 440 test samples, 435 were classified correctly, with 5 misclassifications. Training required approximately 45 seconds, while inference completed in under one millisecond. The resulting model file is 2.5 MB in size, a factor relevant to deployment on mobile or low-power edge devices.

TABLE II: XGBOOST CLASSIFIER PERFORMANCE METRICS

Metric	**Value**
Accuracy	98.86%
Precision (Weighted Average)	0.99
Recall (Weighted Average)	0.99
F1-Score (Weighted Average)	0.99
Cross-Validation Score	97.2% ± 1.3%
Training Time	~45 seconds
Inference Time	<1 ms
Model Size	2.5 MB

F. Profit Prediction using XGBoost Regressor

To capture the economic dimension of each recommendation, a separate XGBoost regressor was trained to estimate expected profit. As the source dataset does not include profit data, synthetic profit values were generated based on plausible yield and price scenarios. This represents a simplification, since real-world profit depends on additional factors such as labor cost, input cost, and market fluctuations. The purpose of this component is to demonstrate that an economic layer can be incorporated into the system, rather than to provide certified financial guidance.

The regression model performed well, achieving an R^2 score of 0.92, indicating that environmental factors accounted for approximately 92% of the variance in synthetic profit. The root mean squared error was approximately 35.13, with a mean absolute percentage error of 4.2%. Cross-validation produced a mean R^2 of 0.89 with a low standard deviation. Training required approximately 38 seconds, with inference completing in under one millisecond.

TABLE IIII: REGRESSION MODEL PERFORMANCE METRICS

Metric	**Value**
Mean Absolute Percentage Error (MAPE)	4.2%
Training Time	~38 seconds
Inference Time	<1 ms
Model Size	2.5 MB

G. Implementation Environment

The system was implemented in Python. Pandas and NumPy were used for data handling, Scikit-learn for scaling and encoding, and the XGBoost library for the classifier and regressor. The autoencoder was implemented using TensorFlow and Keras, with Matplotlib and Seaborn used for visualization. Trained models, the label encoder, and the scaler were serialized after training to enable prediction without retraining, supporting practical deployment scenarios in which low-latency advisory output is required

V. RESULTS AND DISCUSSION

This section presents the experimental results obtained from the proposed Agricultural Decision Support System. The performance of the machine learning models was evaluated using classification and regression metrics to assess their effectiveness in crop recommendation and profit prediction. The system combines climatic analysis, feature engineering, and economic evaluation to generate reliable agricultural recommendations.

A. Crop Recommendation Performance

An XGBoost classifier was trained to predict the most suitable crop from soil and climate inputs. Training accuracy reached 100%, while test accuracy settled at 98.86%. The relatively small gap between these values suggests that the model generalizes beyond memorization of the training data and captures a meaningful relationship between environmental conditions and crop suitability.

To verify that this performance was consistent across different data partitions, 5-fold cross-validation was performed, yielding a mean accuracy of 98.98% with a variance of ±0.46%, indicating that model performance remains stable across repeated training and shuffling of the data.

Precision, recall, and F1 scores were consistently high across most crop categories in the classification report. Rice, maize, coconut, and banana each achieved near-perfect scores, consistent with the strong association between rice cultivation and high rainfall combined with specific temperature ranges, which the model appears to have captured effectively. Comparable strengths in agricultural prediction have been reported in prior work using gradient boosting methods [9].

B. Feature Importance Analysis

A notable advantage of tree-based models is their capacity to report relative feature importance. The XGBoost importance ranking identified rainfall, nitrogen, and temperature as the primary predictive drivers, with soil pH and humidity contributing to a lesser extent. This ranking aligns with established agronomic expectations regarding water availability and soil nutrient profile, while also adding a layer of transparency to the recommendation process. A farmer observing low nitrogen levels alongside a legume recommendation, for instance, can readily interpret the underlying rationale.

C. Profit Prediction Performance

Crop recommendation alone provides only partial decision support; profit estimation was therefore addressed using an XGBoost regressor trained on synthetic profit values derived from plausible yield and price scenarios. Training R^2 reached 0.9931, while test R^2 was 0.8991. This reduction is expected, and reflects the inherently higher noise present in profit prediction relative to crop classification, given the influence of fluctuating market prices, pest incidence, and labor costs. The model effectively learns the relationship between environmental variables and the simulated profit values provided during training.

Predicted profit values ranged from 520 to 4,890, with a mean of approximately 2,150. These figures are plausible but remain dependent on the quality of the synthetic data used for training; a production deployment would require actual historical profit data linked to corresponding soil and climate records

TABLE IIIV: PREDICTED PROFIT STATISTICS

Statistic	Value (₹)
Minimum Predicted Profit	520
Maximum Predicted Profit	4,890
Mean Predicted Profit	2,150
Median Predicted Profit	2,080
Standard Deviation	850

TABLE V: AUTOENCODER RESOURCE USAGE METRICS

Resource Usage	Value
Training Memory	~150 MB RAM
Model Size (Saved)	5 MB (.h5 file)
Encoding Time (440 samples)	12 ms
Decoding Time (440 samples)	15 ms
Total Inference	27 ms (440 samples)
Per-sample Encoding	<0.03 ms
Per-sample Decoding	<0.04 ms

D. Autoencoder Feature Extraction Results

A deep learning autoencoder was used to compress the original seven features into four, with the objective of reducing noise and allowing downstream models to focus on the most informative patterns. The reconstruction error remained low, at 0.1157 on training data and 0.1183 on test data, indicating that essential structural information was preserved.

The resource footprint of the autoencoder was modest: the saved model was 5 MB in size, with encoding of 440 samples requiring approximately 12 milliseconds, or under 0.03 ms per sample. This places the computational cost within range for deployment on modest hardware. Comparable feature extraction techniques have been applied effectively in other agricultural systems [7].

E. Example Crop Recommendation

To demonstrate the functionality of the proposed system, an example prediction was performed using environmental conditions drawn from the dataset: nitrogen approximately 90, phosphorus 42, potassium 43, temperature approximately 20.9°C, humidity 82%, pH 6.5, and rainfall approximately 203 mm. The system returned rice as the recommended crop with high confidence, with a corresponding profit prediction of approximately ₹75,724. This figure, derived from the synthetic profit model, illustrates the type of output presented to the end user: a specific estimated return rather than a crop name alone.

TABLE VI: SAMPLE CROP DATASET WITH RICE LABELS

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.88	82.00	6.50	202.94	rice
85	58	41	21.77	80.32	7.04	226.66	rice
60	55	44	23.00	82.32	7.84	263.96	rice
74	35	40	26.49	80.16	6.98	242.86	rice
78	42	42	20.13	81.60	7.63	262.72	rice
69	37	42	23.06	83.37	7.07	251.06	rice
69	55	38	22.71	82.64	5.70	271.32	rice
94	53	40	20.28	82.89	5.72	241.97	rice
89	54	38	24.52	83.54	6.69	230.45	rice

F. Discussion

The high classification accuracy indicates that environmental data alone provides substantial predictive power for crop recommendation. The XGBoost algorithm, given its capacity to model nonlinear interactions, is well suited to this domain. The autoencoder introduced a layer of dimensionality reduction that likely reduced redundant noise, although the classifier also performed strongly on the raw feature set; the practical benefit of this component may become more apparent on noisier, real-world sensor data

The economic layer represents a more distinctive contribution of this work. Most existing crop recommendation systems address only environmental suitability. Incorporating profit prediction extends the system toward the question of practical relevance to a farmer: expected financial return. The profit model, however, is only as reliable as the data used to train it; as synthetic profit values were used in this study, the corresponding results should be interpreted as a proof of concept rather than a production-grade estimate. A deployed system would require real market prices, input costs, and potentially regional subsidy data.

Interpretability remains an additional consideration. The feature importance analysis and the AI advisory module together provide transparency and natural-language explanation, but user trust develops gradually. Experienced farmers may not adopt a model's recommendation immediately regardless of measured accuracy. While the system is designed to be transparent, adoption is expected to depend on sustained engagement, for example through pilot programs and iterative feedback from end users.

VI. CONCLUSION

This study presented an Agricultural Decision Support System that applies machine learning to recommend crops and forecast profitability based on environmental variables. The core components are an XGBoost classifier, an XGBoost regressor, and an autoencoder used for feature construction. Results under the synthetic conditions evaluated indicate high classification accuracy and a plausible profit estimate.

Several directions are identified for future work. Integrating IoT sensors with real-time weather data would address the limitations of a static model reliant on historical averages, enabling a more responsive system. Satellite imagery could provide an additional layer of information by capturing crop health and soil variability across larger areas.

Incorporating more accurate economic data, including real market prices, input costs, and potentially forward contracts, represents a further avenue for improvement, transforming profit prediction from an estimate into a more actionable financial planning tool.

Finally, deployment as a mobile application capable of operating without continuous internet connectivity would improve accessibility for farmers. The trained models are sufficiently compact for smartphone deployment, and the advisory module could be adapted to function effectively under limited connectivity. These directions represent the next steps planned for this work.

VII. ACKNOWLEDGMENT

The authors would like to thank the Department of Computing Technology, SRM Institute of Science and Technology, for providing the resources and support necessary to carry out this research. The authors also acknowledge the contributors of the CropRecommendation dataset made available on Kaggle, which formed the basis of the experimental evaluation in this study. Sincere thanks are extended to Dr. Usha G for her guidance and supervision throughout the course of this work.

REFERENCES

[1] Prity, Hasan, Saif, et al., “Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations,” Journal of Agricultural Informatics, Springer, 2024.

[2] Shams, Gamel, Talaat, “Enhancing Crop Recommendation Systems with Explainable Artificial Intelligence,” Neural Computing and Applications, Springer, 2023. Available:

[3] Mahale, Khan, Kulkarni, et al., “Crop Recommendation and Forecasting System for Agriculture using Machine Learning,” Discover Sustainability, Springer, 2024.

[4] Sawant, Kumar, Pant, Kotecha “An IoT-Driven Machine Learning System for Smart Crop Recommendation, and optimization in precision agriculture” Journal of Discover Artificial Intelligence, Springer, 2026.

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, August 2026 Impact Factor: 8.2

[5] Sowmya, Meeradevi, et al., “Leveraging Machine Learning for Intelligent Agriculture,” Discover Internet of Things, Springer, 2025.

[6] Abdullahi, Jimale, Ahmed, Nageye., “Revolutionizing Somali agriculture: harnessing machine learning and IoT for optimal crop recommendations,” journal Discover Applied Sciences.

[7] Sudha & Loret, “A review on machine learning-based precision agriculture techniques for crop farming monitoring with IOT”j ournal Discover Environment

[8] Zhang et al., “Design and Implementation of Crop Recommendation Fertilization Decision System Based on WEBGIS at Village Scale” in Advances in Agricultural Informatics, Springer, 2022.

[9] Niamh Creedon, Caoimhe Robinson, Emer Kennedy, Alan O Riordan., “Agriculture 4.0: Development of Seriological on-Farm Immunosensor for Animal Health Applications,” in Proceedings of IEEE International Conference on Smart Agriculture, IEEE, 2019.

[10] Nicoleta Tantalaki, Stavros Souravlas , Manos Roumeliotis, “Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems,”