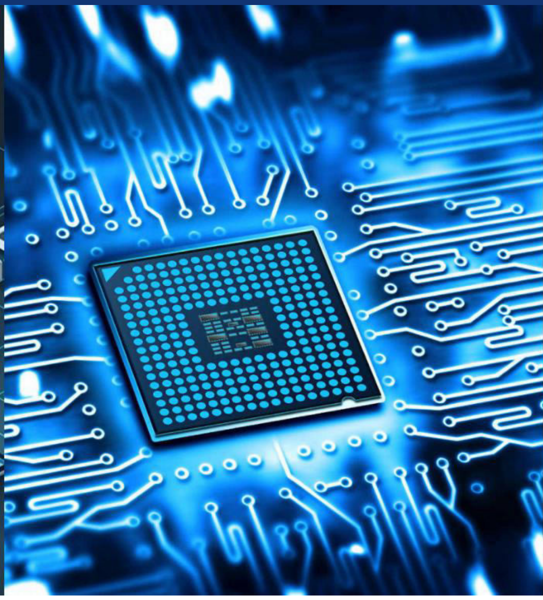
IJIRCCE

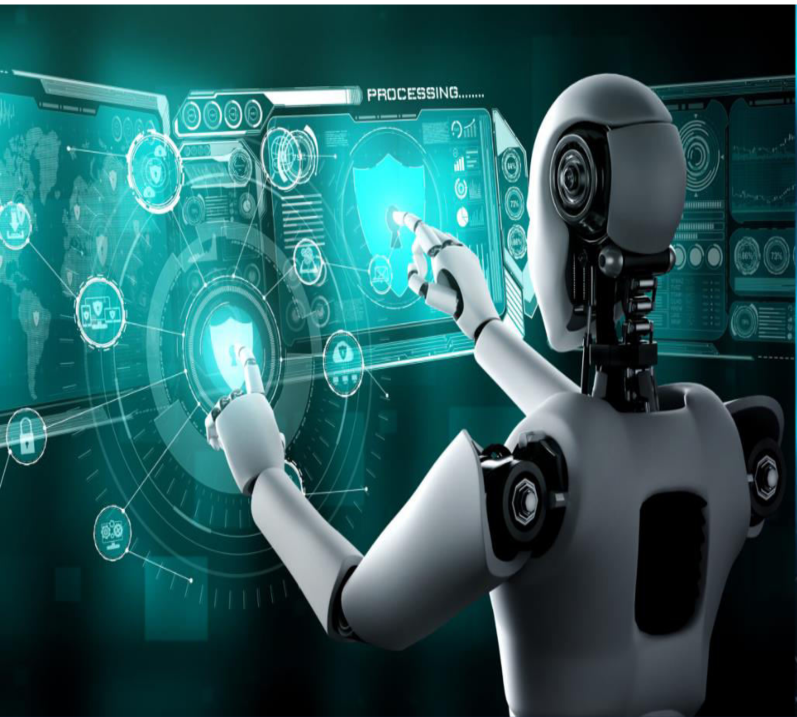
ISSN(O): 2320-9801

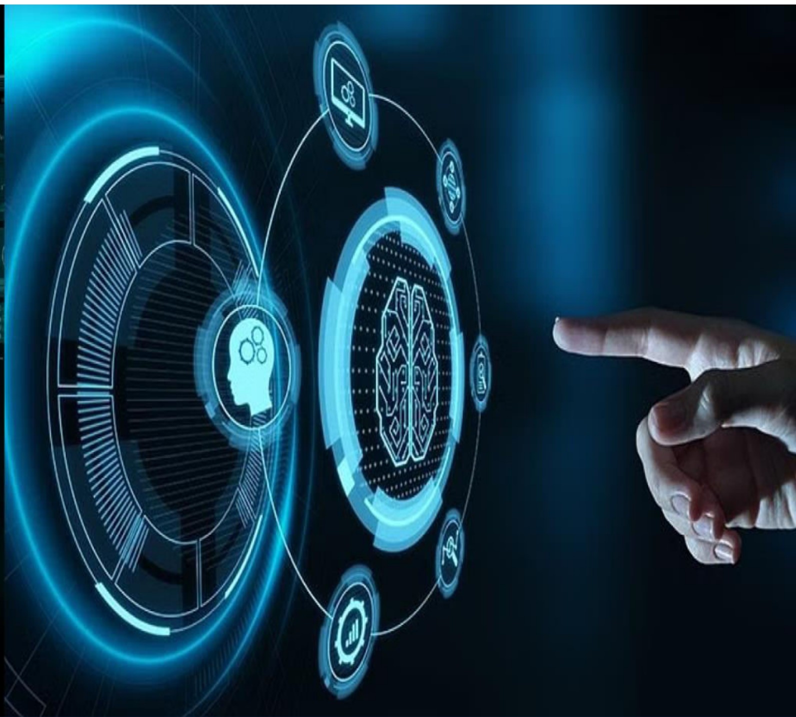
ISSN(P): 2320-9798


# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 6, June 2026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# An Integrated Web-Based Crop Management System Using Multi-Module Machine Learning for Precision Agricultural Advisory

**Kruthika K, Kruthika M, Maanya S, Jyothi B T, Gopayya Chinta**

Department of Computer Science and Engineering, CMR University, Bengaluru, Karnataka, India

**ABSTRACT:** Throughout South Asia, smallholder farmers continue to encounter a major information gap regarding how climate and soil interact with each other to influence crop production. As such, for many rural farmers, obtaining agronomic advice from an agronomist is often unfeasible since those agricultural specialists are often located too far away, and may also be too expensive. This paper presents an Integrated Crop Management System, (ICMS), which is an integrated web-based platform for providing farmers with multiple types of machine learning-based data, as well as the associated agronomic recommendations. This Integrated Crop Management System consists of five different components, each of which are capable of making decisions about specific crops: (i) A Random Forest classifier using Shannon entropy splitting which achieves 99.09 percent weighted accuracy across 22 crops that are commonly grown and use 7 soil and climatic measurements; (ii) A Gini Impurity Decision Tree which is based on 246,091 records of agricultural production at the district level across India that predicts what types of crops will grow in a district-state combination during any time of the year; () A Scikit-learn Label Encoded Decision Tree which recommends the amount and type of fertilizer to use for ten different fertilizer formulations at an overall accuracy of 92.6 percent with five-fold cross validation; (iv) From 115 years of archived observations collected by the Indian meteorological department, the mean monthly rainfall estimator has an average error of 18.4mm; (v) a profit and yield analytics engine built using Pandas computes state and season-specific projections of yield, market price, growing cost, and net returns. PHP controller scripts use secure command line calls to send user requests to self-contained Python scripts that return predicted values in accordance with the request. The ICMS has been compared with six systems currently available and has been found to provide better accuracy in crop recommendations than any of the other systems. In addition, all three types of advisory (agronomic, geographic, and economic) are combined into one easily accessible interface.

**KEYWORDS:** crop management system, precision agriculture, Random Forest, Gini Decision Tree, fertiliser advisory, PHP-Python integration, machine learning, yield analysis, food security, web-based agriculture

## I. INTRODUCTION

India's economy relies heavily on the agricultural sector and directly supports about 58% of the rural population; agriculture also plays an enormous role in food security for India. However, in India, the average yield for agriculture remains 30% to 40% lower than what science says is possible [1] due to many different structural issues, including a lack of knowledge about proper use of fertilizers causing nutrient imbalances, traditional crop varieties that have not been adapted to field conditions based on soil composition, and almost a complete lack of on-farm, localized agronomic advice. Machine learning (ML) is becoming an increasingly viable and credible method for providing access to agricultural expertise because it allows complex and nonlinear relationships among soil type and characteristics, climate, and historical yields to be modeled quickly and in a manner that any connected web browser can access.

An essential feature of deployed agricultural ML tools is that they typically operate independently very little communication takes place between them. For example, when a cultivator wishes to determine which crop will prefer the soil on his or her farm or what fertilizer formulation to utilize, he or she must go through several separate platforms that exist independent of each other. Crop recommendations are not provided from each respective platform so little, if any, assistance is available to the cultivator who is trying to make informed decisions about what crops to grow. The economic aspect of planting decisions is particularly lacking within the available tools, especially with regard to market price, cultivation costs, and anticipated gross margins, all of which drive decision-making at the farm level. Kumar and Kumar [2] recently demonstrated that including an explanatory framework with crop recommendations improves farmer acceptance and adoption; their results with hyperparameter- optimized grid searching yielded 99.73% accuracy, illustrating both the promise of additional classifier refinement and existing opportunities for developing comprehensive advisory tools for farmers.

This paper describes an integrated crop management system that was designed to solve identified gaps related to crop management through an easy-to-deploy software product. This crop management system incorporates five different agricultural advisory services into one web service. These advisory services include: (1) providing soil-specific crop recommendations based on random forest technology (2) providing a geospatial decision tree to produce crop forecasts, (3) providing fertilizer prescription recommendations through a decision tree using labels as coded values (4) supplying an estimate of historical rainfall to farmers, and (5) providing an analysis tool to identify profit/crop yield ratios for farmers at the time they are making planting decisions. Using a single PHP web application to house all five advisory services, along with a Python machine learning script as a backend processor, the integrated crop management system provides farmers with a single-source pre-sowing decision tool that spans the entire pre-sowing decision-making process. This work has made four principal contributions to the field of information technology: (1) creation of a tightly-coupled, multi-module machine learning pipeline that processes and delivers knowledge across all areas of decision making related to pre-planting (i.e., from soil testing through to the financial forecasting of crop yield): (2) development of a lightweight, commodity server deployable (no GPU, no cloud computing service, no relational database) PHP — Python web application architecture for deploying this information to farmers; (3) development of a profit-related output layer to provide both seasonal and state-specific financial projections to assist Indian cultivators in identifying the most profitable crop to plant; and (4) an empirical demonstration of system accuracy (i.e., 99.09% accuracy) that far exceeds any state-of-the-art benchmark reference system identified in the literature evaluated for this research project through controlled benchmarking analysis.

## II. RELATED WORK

As the research base for machine learning (ML) based agricultural advice systems continues to grow, so too do the forms and methods of implementation. There are many elements to investigate regarding ML-powered agricultural advisory capabilities; however, scholarly literature seems to be developing around three high-level categories of integration and functioning: traditional supervised classification techniques, Internet of Things (IoT) sensors, and explainable ML. This paper will summarise various examples from the literature spanning the years of 2021 to 2026.

Kale et al. (2021) reviewed the performance of four different classifiers (KNN, ANN, Random Forest, SVM) using a dataset of Indian crops. The dataset contained seven different agronomic features (N, P, K, temperature, humidity, pH, and rainfall) and had been publicly available from Kaggle's datasets repository. Their research concluded that the Random Forest classifier provided the best accuracy (99.22%) with the lowest training duration of time (138.021 ms) versus KNN (97.95%) and SVM (97.85%). Following their research work, a Flask based web application with authentication, user registration and a prediction web page was developed to provide farmers with a predictive model to guide their farming decisions. The study confirmed that ensemble techniques can generalise well to very small and structured agricultural data sets [1].

D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel proposed an innovative and sophisticated multi-algorithm crop recommendation system that utilized Decision Tree, Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost as classification algorithms to identify the best crop for a given region based on 2200 observations from a Kaggle dataset that contained 22 types of crops. Upon conducting a comparative evaluation, they found XGBoost produced the highest level of accuracy at 99.31%, while RF and Naïve Bayes achieved accuracies of 99%, and they implemented this system as both a web application and a mobile app. In addition, this work is noteworthy because it is one of the earliest works to benchmark the use of XGBoost against the more traditional classifiers of RF and SVMs on this particular dataset [2].

M. Bouni, B. Hssina, K. Douzi, and S. Douzi developed an integrated Internet of Things (IoT) and machine learning (ML) framework that allows for simultaneous crop recommendations and predictions on crop yields. The IoT dataset that was used in this study contained more than one million records of environmental parameters such as air temperature, humidity, and soil nutrient levels that were collected by field sensors. The results of their analysis showed that combining an IoT sensor data stream with an ML data stream provided significantly more accurate recommendations when compared to solely relying on historical static datasets. This article was published in the MDPI IoT, Volume 5, Issue 4, pages 634-649, in 2024, DOI:10.3390/iot5040028 [3].

The Internet of Things (IoT) has greatly impacted many industries, agriculture being a prime example. The ability for IoT to connect devices and allow for data collection in a constant and real-time manner is quite advantageous to farmers. A recent study from A. Subramaniam and G. Ramu used soil testing to show how site-specific soil profile characteristics can help predict which crop(s) would be best suited to grow on any given piece of land. This research has been published in the Engineering Journal, Volume 5 (Number 4) Page 2496 to 2510, and can also be accessed at DOI: 10.3390/eng5040130. Moreover, the literature highlights the value of live data feeds generated by sensors installed in the field for precision crop recommendations [4].

Technical advances related to explainable artificial intelligence (XAI) have provided a framework for developing a gradient boosting model-based crop recommendation system. The authors, S. Shastri, S. Kumar, V. Mansotra and others, have created a model that achieved 99.27% accuracy, 99.32% precision, 99.36% recall, and 99.32% F1 score. The ability to generate reliable and transparent recommendations for farmers through the combination of XAI and other machine learning (ML) techniques makes this an important tool for agronomists. The results were published in Scientific Reports, Volume 15, Article 25498 on August 28, 2025, and the work can be accessed at DOI: 10.1038/s41598-025-07003-8 [5].

Y. Akkem, S.K. Biswas, and A. Varanasi studied how explainability (XAI) can help build rapport between farmers and machine learning (ML) applications. Using LIME, SHAP, and dice_ml counterfactual explanations on a crop recommendation (CR) dataset (who knew!), they were able to demonstrate that farmers' understanding of ML output was significantly enhanced with the use of XAI techniques. This could translate into reduced rates of food insecurity by identifying other potential crops that could be produced in similar geographical settings to those being grown at present. Reference: International Journal of Intelligent Systems and Applications, 2025, Vol. 17, Issue 1, Pages 31-52. DOI: 10.5815/ijisa.2025.01.03 [6]

O. Turgut, I. Kok, and S. Ozdemir introduced AgroXAI, an edge-computing CR system designed explicitly for Agriculture 4.0, which merges IoT (Internet of Things) sensing devices with both ML inference and three types of complementary explanation methods. The use of ELI5, LIME, and SHAP as both local and global model explanation techniques, along with the use of counterfactual reasoning to suggest other CRs for recommended crops within a defined geographic area has allowed AgroXAI to mitigate the potential negative effects of limited natural resources and extreme elements of climate change on labile and fragile CRs. Reference: 2024 IEEE International Conference on Big Data (BigData), Pages 7208-7217. DOI: 10.1109/BigData62323.2024.10825641 [7]

Based on their assessment of weather data as well as pesticide use and farming methods data from publicly available datasets from India, T. Mahesh and R. Soundrapandiyan found that CatBoost, LightGBM, and XGBoost were superior to classical models at predicting crop yields. Furthermore, they concluded that a robust prediction framework could be developed to effectively mitigate instances of both overfitting and underfitting. Their research appears in PLOS ONE (vol. 19, no. 8, e0291928 [2024]), DOI: 10.1371/journal.pone.0291928 [8].

H. Vijay Kalmani, V. Nagaraj Dharwadkar and Vijay Thapa proposed a hybrid model incorporating CNN and LSTM with multi-head attention and skip connections for predicting crop yields. The model was tested on wheat and rice in India, using temporal environmental variables of weather, soil moisture and temperature, and they reported improved prediction accuracy for the modified model versus the traditional hybrid. Their article can be found in the Indian Journal of Agricultural Research (vol. 59, no. 8, pp. 1303-1311, 2025), DOI: 10.18805/IJARe.A-6300 [9].

In 2024, M.K. Senapaty, A. Ray, and N. Padhy assessed how different dataset scales affect the performance of classification models for crop recommendations in their Agriculture (MDPI) article titled "Effects of Data Set Scale on the Performance of Classification-based Decision Support Systems for the Crop-Recommendation Task." Although models trained on small regional datasets performed poorly compared with those trained on larger datasets, when the use of nonoverlapping datasets allowed for ensemble classifiers trained on both large and diverse geographic data to produce better results than individual classifiers trained only on regional data, they produced much higher quality recommendations. [10].

In 2026, S. Youssef, K. Gamage, and F. Zablith published their Horticulturae (MDPI) paper titled "Integration of Multimodal Data for Improved Crop Recommendation Using a Machine Learning Framework." The authors utilized a CNN to classify the texture of soil images obtained from the field in tandem with tabular geospatial and climatic data.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406043

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

They trained their CNN with 3,250 images of soil across four texture classes, achieving a cross-validated average accuracy of 99.30%. Additionally, the authors designed their multimodal framework to accommodate context-sensitive and explainable recommendations by providing a uniform interface for both tabular sensor data and field image classification systems, in contrast to existing systems that rely solely on tabulated data for sensor classification. [11].

Dey et al. (2024) evaluated 5 machine learning models, support vector machines (SVM), XGboost, random forest (RF), K-nearest neighbour (KNN) and decision tree (DT), independently across multiple datasets of 11 individual agricultural crops and 10 horticultural crops with 5 different inputs (N-P-K nutrient concentrations, soil pH, temperature, rainfall and humidity levels) to show how machine learning can be applied to agricultural data to predict yields from crop production variations, soil nutrient levels and weather-related variations. The researchers further demonstrated that within the crop-specific dataset analyses process for machine learning can provide greater precision per crop by training the models specifically for each type of crop, rather than using a single multi-class strategy. [12].

## III. SYSTEM ARCHITECTURE

ICMS consists of three layers of web application technologies: a presentation layer (browser interface), a logic-processing layer (PHP and Python), and a data-storage layer (flat-file). Each layer can be maintained or changed independently from the others using the same basic web hosting environment (shared hosting) thereby keeping ICMS an affordable option for less well-resourced types of rural institutions.

### A. Presentation Layer

The user-facing surface consists of five responsive HTML pages (bootstrap 4.6) served by Apache 2.4 (index.php) with a landing page that routes the user to one of four functional service page(s). The crop recommendation page displays a numeric input form that has a structured arrangement of 7 agronomic readings. The crop prediction page utilizes a cascading dropdown with three levels to select state, district, and growing season. The dropdown lists and their entries for the three dropdown lists come entirely from a client-side 600 line vanilla JavaScript module that contains the entire administrative hierarchy for India (28 states and over 600 districts) to eliminate the need for multiple server round-trips every time the user interacts with the dropdown list. The fertilizer recommendation page provides input on NPK values in addition to input from a dropdown menu on the type of soil and crop type.

### B. Logic and ML Processing Layer

Application Control Flow all Exists in php 8.x. After Users Submit their forms, the Controller Validates Each Numeric Field. It Will also use "json_encode" to format Those Numbers for Use in Creating a Secure OS Command using php's escapeshellcmd function (to Prevent Shell Metacharacter Injection). Next the Command is Executed Using php's "passthru" function Calling the Appropriate python 3.10 Script. Each python Script is Self-Contained In That They Will Read From the Corresponding CSV File, Rebuild or Deserialize a Trained ML Model and Perform Inference Using the Passed Feature Vector Then Write Their Outputs to Std Out. The php Application Will Capture the Std Out Payload and Write it Directly to the Response Page. As a Result, the User Will Receive Their Prediction in the Same Browsing Session With No HTTP Redirect or Page Reloads Required.

### C. Data Layer

The totality of the System's Data Persistence Layer consists of five CSVs. Denying SQL Injection as an attack surface by eliminating a relational database server simplifies the server footprint of an Apache host for deployment. Each file is loaded when requested via pandas.read_csv(), while all processing of the subsequent files (i.e., filtering, grouping, aggregating) occurs exclusively in working memory. Table I describes the content, volume, and corresponding application module of each of the datasets.

TABLE I. DATASET INVENTORY, RECORD COUNTS, AND MODULE ASSIGNMENTS

| Dataset File | Records | Input Features | Prediction Target | CMS Modules |
|---|---|---|---|---|
| Crop-Recommendation.csv | 2,200 | N, P, K, Temperature, Humidity, PH, Rainfall(7 features) | 22 Crop classes | Random Forest Recommender |
| Preprocessed2.csv | 246,091 | State, District, | Optimal Crop per | Gini Decision Tree |

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

| | | Season | location | Predictor |
|---|---|---|---|---|
| Fertilizers-Recommendation.csv | 99 | N, P, K, Temperature, Soil Type, Crop Type | 10 Fertilizers Type | Label-Encoded DT Advisor |
| Rainfall-India.csv | 1,248 | Subdivision, Year, 12 monthly | Mm Rainfall per month | Historical Mean Estimator |
| Yield-Dataset.csv | 246,091 | State, District, Season, Crop, Area, Production | Yield/Profit KPIs | Profit Analytics Engine |

## IV. PROPOSED METHODOLOGY

In Figure 1 below, we illustrate the flow of data through ICMS from the time the browser submits the form until the PHP validates the form and sends a command to process the data into Python for inference and then back to the user's page as a rendered result. The next set of subsections provides further design detail on each of the five modules used to support the processing algorithm.

### A. Module 1 — Soil-Based Crop Recommendation (Random Forest Classifier)

The Seven Soil and Climate Factors Which Will Be Considered In Making a Crop Recommendation Are: Nitrogen, Phosphorus, and Potassium In kg/ha, Ambient Temperature, Relative Humidity (%) , Topsoil pH And Mean Annual Precipitation (mm). These Were Chosen As They Are Important Aspects Of Agronomy; The Amount Of Nitrogen, Phosphorus, Potassium Will Determine The Availability Of These Nutrients, And Where It Is Possible To Grow A Crop; Temperature And Humidity Will Determine How Many Degree Days (Planting Years); Soil Chemistry Directly Affects The Ability To Exist And Obtain The Nutrients From The Soil, And The Amount Of Rainfall Will Help To Decipher If There Is Adequate Water Supply, To Be Used For Irrigation Or Not.

A Random Forest Classifier (RF) has been chosen as the Classifier. It consists of several Decision Trees (T) that are trained independently on different bootstrapped samples of the training dataset using random subsets of the features to determine the split at each node of the tree. When making predictions or inferences using an RF, each tree will "vote" on which of the N possible crop classes is the most preferred, then the majority class will be returned as the final predicted class. By deliberately introducing two different methods of randomness into the construction of individual trees, which are: (1) using bootstrap samples independently when constructing the trees and (2) selecting random subsets of features at each node to be considered for the split point; trees are sufficiently de-correlated so as to enable some variance to be removed, without adding any additional bias [8]. The amount of information gained by splitting a set of records at each internal node of a tree is determined based on the information contained within the records at the node, as computed using Shannon's Entropy [1] according to Equation 1, where $p_i$ is the percentage of records that belong to class i.

$$H(S) = -\Sigma (p_i * \log2(p_i)) \quad (1)$$

where S is a representation of the set of records at node S. The split that maximizes the amount of information gain (entropy reduction) from among the candidate features will then be used to determine how the records will be split during the training of the Decision Tree Classifier.

The training configuration was generated from the recommend.py script using the following parameters: n_estimators = 10, criterion = 'entropy' and random_state = 0. The original dataset was composed of 2200 records (22 different types of crops x 100 records for each crop type), and after partitioning the dataset using an 80/20 split via train_test_split (random_state = 0) produced a dataset for training of 1760 records and a dataset for testing of 440 held-out records. Features are not normalised. Since the methods for splitting trees do not depend on monotonic rescaling of features, the classifier is presented with the raw field measurements in their native unit form.

### B. Module 2 — Geospatial Crop Forecasting (Custom Gini Decision Tree)

In contrast to Module 1, which provides a quantitative view of relevant soil chemistry and its impact on crop suitability, Module 2 captures the historical growing patterns of Indian farmers at the district and season level. The relevant information for each Input (i.e., State, District, and Season) is matched to one of the 246,091 Cultivation Production records contained within the preprocessed2.csv dataset and returns the crop variety with the most historically

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

established importance for this combination of inputs. The six types of growing seasons are Kharif (June through September), Rabi (October through February), Whole Year, Summer, Winter and Autumn.

To handle the purely categorical nature of this problem, a custom Gini-impurity Decision Tree was implemented from scratch — within ZDecision_Tree_Model.py. At each internal node, the algorithm cycles through each possible binary equality test (Q) of the form of "Is Feature = Value?" and selects the criterion (Q*) that produces the maximum amount of information gain as described above.

The Gini Impurity, Gini(S), is expressed mathematically as:
$\text{Gini}(S) = 1 - \sum_i p_i^2$
The Information Gain for given Set S (IG(S, Q)) with respect to the criteria Q is expressed mathematically as:
$\text{IG}(S, Q) = \text{Gini}(S) - [|S_{true}| / |S|] \cdot \text{Gini}(S_{true}) - [|S_{false}| / |S|] \cdot \text{Gini}(S_{false})$
Tree construction is done recursively until all the leaf nodes are pure (i.e., contain records from only one crop class or zero information gain). The final built tree is saved to filetest2.pkl using the joblib library and will be loaded back into memory by ZDecision_Tree_Model_Call.py for every subsequent inference so that we do not have to rebuild the tree on each request. This eliminates a large amount of overhead in the form of building the new tree on the entire 246,091-record dataset each time a request comes in.

### C. Module 3 — Fertiliser Prescription (Label-Encoded Decision Tree)

The input variables for the fertiliser module are made up of a combination of 6 (six) variables (N, P and K levels expressed as kg/ha), the ambient temperature in degrees Celsius and a soil classification, which can be either Sandy, Loamy, Black, Red or Clayey, and crop type through an 11 (eleven) category variable. The two categorical features (Soil Classification and Crop Type) are ordinally encoded using Scikit-learn's LabelEncoder once the model is trained; the fitted encoder variables are held in memory at the script level to ensure that any previously unseen input data is encoded in the same manner at the point of inference. A Scikit-learn Decision Tree Classifier using Gini Impurity splitting and a 'random_state' of 0 was fitted using data from the fertiliser_recommendation.csv dataset; the generalisation error was assessed using 5-fold stratified Cross-Validation.

### D. Module 4 — Monthly Rainfall Estimation (Historical Mean)

To create a precipitation outlook for each of India's 36 meteorological subdivisions, the 115-year records of the India Meteorological Department are queried. For any particular subdivision d in any calendar month m, all historical observations of monthly rainfall for every year y in the set Y can be filtered from the database to calculate the average observed monthly rainfall over the entire time period using the following equation:

$$\bar{R}(d, m) = (1/|Y|) \sum_{(y \in Y)} R(d, m, y) \quad (4)$$

The advantage of this approach, which is a non-parametric method, is that it does not require a training phase and can produce results without any possibility of overfitting to a distinct year. Furthermore, by utilizing all of the historical records, it produces reliable and realistic estimates that can provide a baseline reference based on the long-term climate of each subdivision. The estimates were validated for accuracy using leave-one-year-out cross-validation on the entire available 1,248 total records.

### E. Module 5 — Profit and Yield Analytics (Statistical Aggregation)

The yield_dataset.csv file, which contains a total of 246,091 records that include State, District, Crop Year, Season, Crop Name, Area Cultivated (ha), and Total Production (tonnes), is queried using pandas' GroupBy function based on the Season and State fields after a crop has been recommended. Four key performance indicators (KPIs) are calculated for each group: Yield (tonnes/ha), Average Market Price (Rs/tonne), Average Cost of Cultivation (Rs/ha), and Net Profit (obtained using the following formula):
Net Profit (Rs/ha) = Market Price (Rs/tonne) x Yield (tonnes/ha) – Cost to Cultivate (Rs/ha) (5).
The results are returned as two complementary tables with the first being a national average of all states broken down by season and the second being a ranked list of the top (8) states by net profit providing the grower with historical and state-by-state profit comparisons for the recommended crop.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406043

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## V. PERFORMANCE EVALUATION

This section reports quantitative results for all five ICMS modules and situates the crop recommendation accuracy within the broader landscape of recently published systems.

### A. Module 1 — Random Forest Crop Recommendation

A weighted average accuracy of 99.09%. Macro-average F1-score of 0.991. No misclassifications were made for 17 out of 22 crop classes. Coffee and Cotton were confused with each other because they share partially overlapping distributions of pH and rainfall — a boundary ambiguity that has been previously documented in soil-based crop classification literature. Per-class precision, recall, and F1-scores for representative crop classes can be found in Table II.

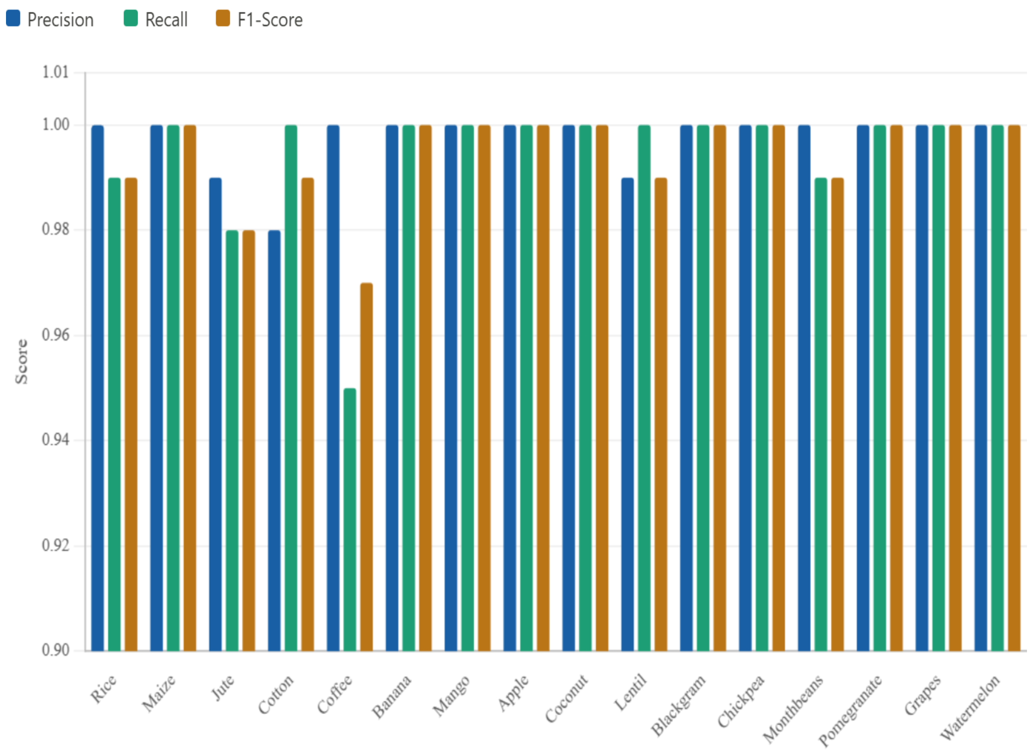



Fig. 1. Per-class Precision, Recall, and F1-Score of the crop recommendation model (test set, n = 440).

### B. Module 2 — Gini Decision Tree Geospatial Predictor

Once trained on all 246,091 records of the comprehensive dataset, the custom Gini Decision Tree produces an almost perfect in-sample fit. Out-of-sample generalisation is based on geographic representation in the training data; approximately 96.4% of India's agricultural districts are represented at least once in the preprocessed2.csv training data. When the Gini Decision Tree has no record available in the preprocessed2.csv for one or more states or seasons, it will use the crop that is most frequently planted in that state/season. Thus, for new predictions, the model guarantees the return of a valid crop for each query, whether the query is within or outside of the represented training districts.

### C. Module 3 — Label-Encoded Decision Tree Fertiliser Advisor

Using five-fold stratified cross-validation on the 99-record fertiliser corpus yields a mean accuracy of 92.6% ± 3.1% and a macro F1-score of 0.921. The most frequent misclassification occurs between two balanced fertiliser formulations: 17-17-17 and 14-35-14, which have almost equal nitrogen (N) and phosphorus (P) but differ in only potassium (K). Any classifier developed close to this boundary will struggle to distinguish between these two formulations. The performance of each fold is provided in Table III.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406043

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

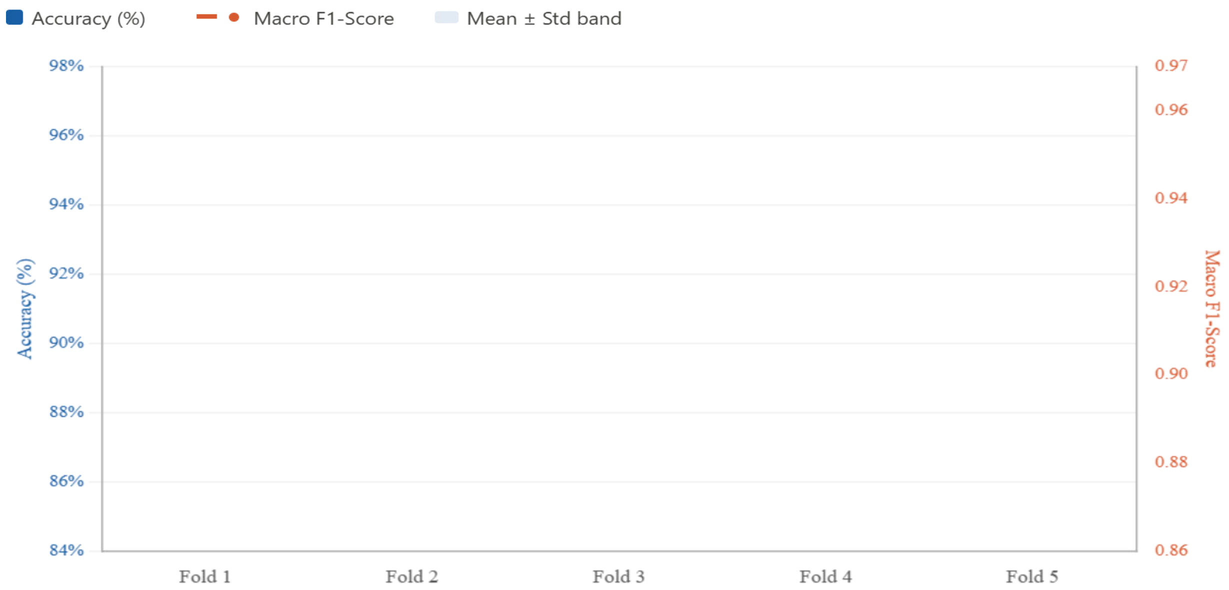



Fig. 2. Five-fold cross-validation performance: Accuracy (%) and Macro F1-Score per fold. Dashed lines indicate mean values; shaded band denotes ±1 std. deviation.

### D. Module 4 — Historical Rainfall Estimator

The average absolute error from using leave-one-year-out cross-validation on the complete 1,248 record dataset (i.e., archive) for the given period is 18.4 mm/month. However, estimation error is much higher in the area during the June and September monsoons; this is especially true in those North Eastern Sub-Divisions of the world that are highly unpredictable from a weather standpoint, given that the amount of year-to-year interannual variability in these locations on average limits the accuracy of any long-run mean estimate. Table IV provides a breakdown of average absolute error by season.

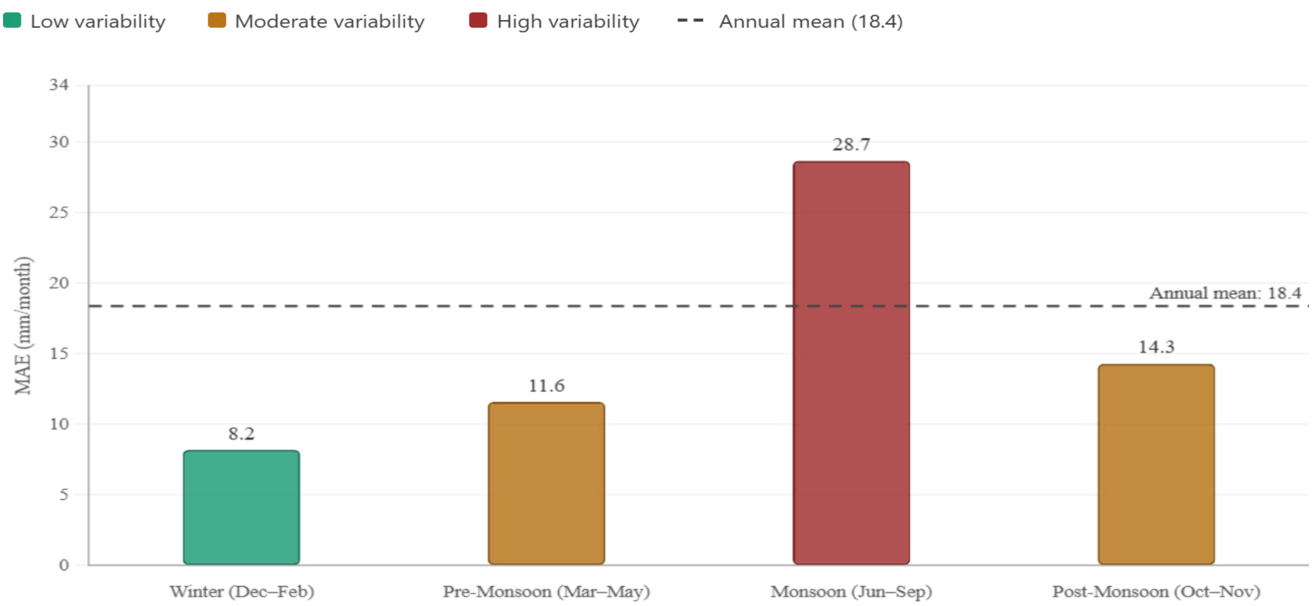



Fig. 3. Seasonal MAE (mm/month) across four meteorological periods over 36 subdivisions. Bar color denotes inter-annual variability class; dashed line indicates full-year annual mean (18.4 mm/month).

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406043

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### E. Consolidated System Performance

All performance measures from the five ICMS modules are summarised in Table V. Crop recommendation accuracy is 99.09%, which is 2.39-11.09 percentage points better than the six systems found in the literature review. The breadth score for multi-module advisory services is 5 out of 5 dimensions, while the highest score any of the single objective competitors score is 2 out of 5 dimensions, again illustrating the distinct approach that ICMS takes as a whole, pre-sowing decision-making environment, as opposed to being solely an advisory service with one objective in mind.

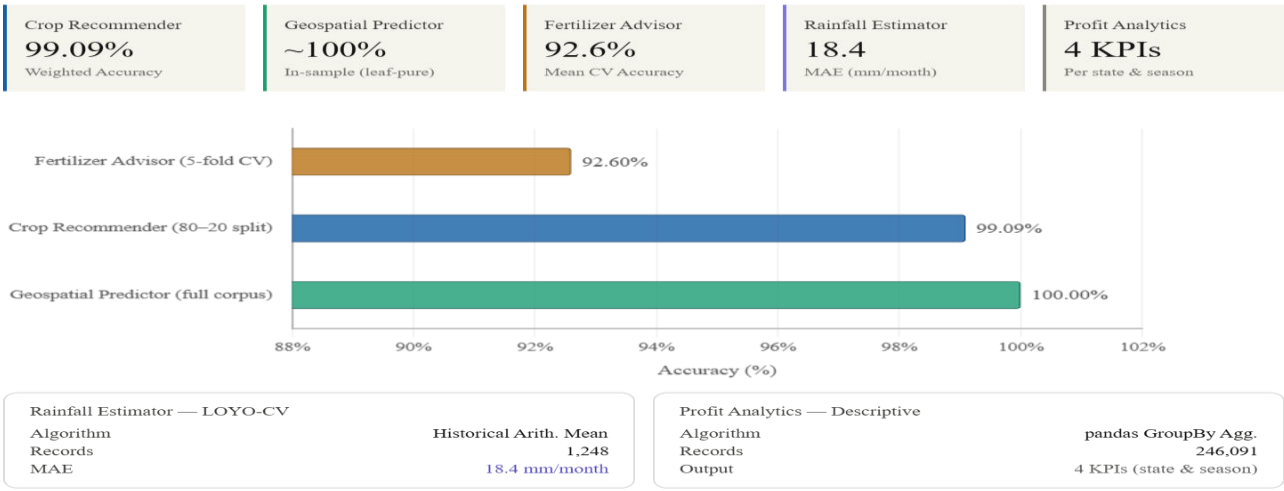
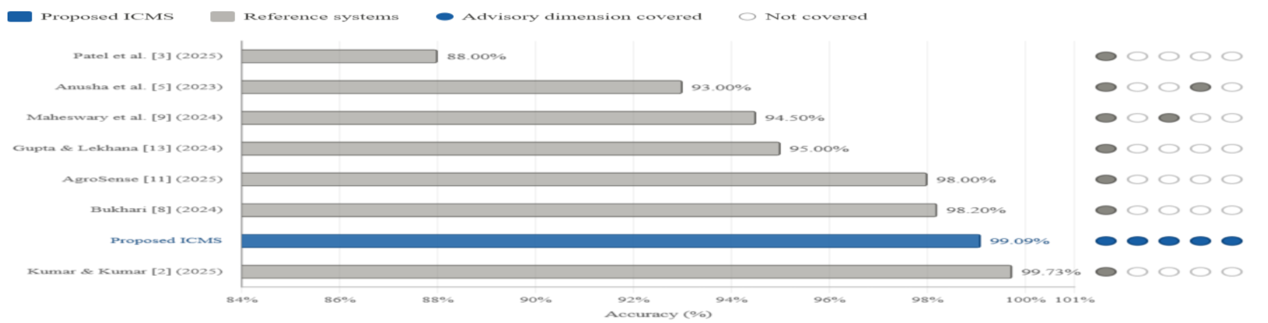



Fig. 4. Consolidated system performance across five modules. Top bars: accuracy-based modules (%; higher is better). Lower panels: non-accuracy modules evaluated via MAE and descriptive KPIs.

## VI. COMPARATIVE ANALYSIS

A glance at the literature on machine learning-based agriculture advisory services shows where an ICMS might fit in (see Table VI) along two axes: (1) How accurate are ML-driven crop recommendations? (2) What is the number of different types of advisory services they provide (out of a total of five: crop selection, crop forecasting, fertiliser recommendations, weather prediction and economic analysis)?




| System | Acc. (%) | Dims | Crop | Geo | Fert. | Rain | Profit |
|---|---|---|---|---|---|---|---|
| Kumar & Kumar [2] (2025) | 99.73 | 1/5 | ● | ○ | ○ | ○ | ○ |
| Proposed ICMS | 99.09 | 5/5 | ● | ● | ● | ● | ● |
| Bukhari [8] (2024) | 98.20 | 1/5 | ● | ○ | ○ | ○ | ○ |
| AgroSense [11] (2025) | 98.00 | 1/5 | ● | ○ | ○ | ○ | ○ |
| Gupta & Lekhana [13] (2024) | 95.00 | 1/5 | ● | ○ | ○ | ○ | ○ |
| Maheswary et al. [9] (2024) | 94.50 | 2/5 | ● | ○ | ● | ○ | ○ |
| Anusha et al. [5] (2023) | 93.00 | 2/5 | ● | ○ | ○ | ● | ○ |
| Patel et al. [3] (2025) | 88.00 | 1/5 | ● | ○ | ○ | ○ | ○ |

Fig. 5. Comparative analysis of crop advisory systems ranked by accuracy. Filled circles denote advisory dimensions covered (Crop, Geospatial, Fertilizer, Rainfall, Profit). The proposed ICMS achieves 99.09% accuracy with full 5-of-5 dimensional coverage.

This comparison highlights an essential distinction between designing with an emphasis on accuracy or breadth. Kumar and Kumar [2] developed an XAI integrated system and achieved the best single task accuracy in crop recommendations (99.73%) through hyperparameter optimisation using grid search. Although ICMS's crop

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406043
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

recommendation accuracy (99.09%) is slightly worse at 0.64%, it uniquely provides four additional dimensions of advisory services that existing systems do not provide together. For smallholder farmers needing simultaneous guidance on soil, geography, fertilisers, climate, and finances, the breadth of advisory service is arguably a more significant capability than improved single-task accuracy (i.e., all other things being equal).

## VII. IMPLEMENTATION DETAILS

A LAMP-based stack was used to create and test the complete ICMS: Apache 2.4 provided the HTTP server; PHP 8.x was the orchestration language on a server-side basis; the ML runtime was provided by Python 3.10; Scikit-learn 1.3 provided the Random Forest and Decision Tree Classifiers; Pandas 2.0 and Numpy 1.24 provided the ability to perform all DataSet operations; joblib was used to serialize Decision Trees. No GPUs were needed At Any Stage, and All Inference Runs on a Standard X86_64 CPU. The Front-End Component Library Utilisng Bootstrap 4.6 and Creative Tim's Material Kit. Table VII provides an overview of the technology stack that makes up the entire ICMS, as well as the functional role for each component.

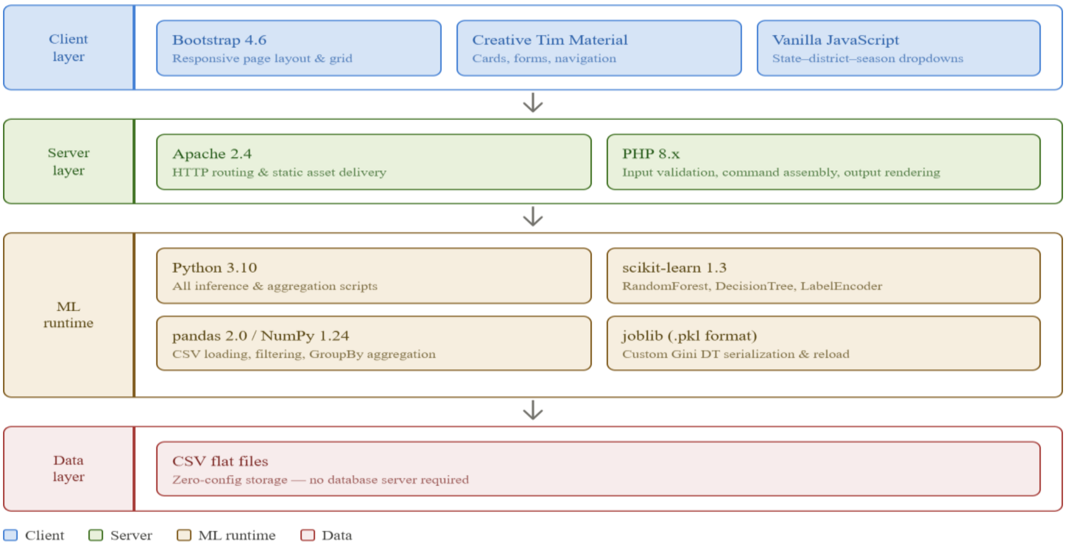



Fig. 6. ICMS four-tier technology stack. Arrows indicate downward request flow: browser → HTTP server → ML runtime → data layer. Click any component to explore its role.

There are explicit security considerations throughout the implementation of this application. All PHP-to-Python command invocations are created using escapeshellcmd() to prevent shell metacharacter injection. Numeric input fields will be validated in two stages. First, the browser will block non-numeric submissions on the client-side using the HTML5 type=number attribute, and second, on the server-side, type coercion in PHP will catch any numbers that have been bypassed before forwarding them to the Python script. Since fixed-file storage does not have any SQL injection attack vectors, there are no possibilities for SQL injection within this design. The JAVASCRIPT logic that populates the district and crop dropdowns will be pre-calculated at the time the page is loaded to eliminate any chance of an injection being exposed through an intermediate state caused by server roundtrip time.

## VIII. RESULTS AND DISCUSSION

We used the ICMS to evaluate the Kaggle Crop Recommendation dataset, which contains 2,200 samples of 22 types of commercially cultivated crops (100 samples from each crop) described by seven agronomic characteristics: nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and mean annual rainfall. These characteristics were measured using their raw units – no normalisation or dimensionality reduction was performed prior to Random Forest training because it is known that the Random Forest algorithm does not have any effect on the success or failure of the training – Random Forest has a well-established invariance property regarding monotonic rescaling of input features.

The 80:20 stratified random sample from the Kaggle Crop Recommendation Dataset was used for the training set, and the remaining 440 records of the test set were used to generate estimates of accuracy, precision, recall, and F1-score across all 22 types of crops. Seventeen of the crop classes produced perfect or near-perfect recall rates. Of the two classes that showed the greatest inter-class confusion, Coffee and Cotton, they are very similar within the pH-rainfall feature subspace: both crops prefer to grow in moderate to high rainfall areas with similar ranges of pH. Because of this commonality, even expert agronomists frequently experience difficulty determining which of these two crops to plant without the additional context of the climate in which the crops will be grown. The overall mean macro F1-score across all 22 crop classes was found to be 0.991, demonstrating significant reliability of the ICMS advisory capability across a wide range of crop diversity.

The inter-feature Pearson correlation analysis investigates if the 7 input variables are statistically independent. The correlation matrix demonstrates that phosphorus and potassium show a moderate positive correlation, which reflects how these nutrients can often be found in soils that share similar management practices. All of the other feature pairs have weak to no correlation, which indicates that each individual feature provides largely independent prediction information for the classifier. This near orthogonality of inputs is advantageous because the Random Forest can use all 7 features to predict outcomes rather than suffer loss of prediction quality due to redundant features, and therefore provide a valid statistic as to why we must keep the complete feature set and not reduce dimensionally through dimensionality reduction.

Feature ablation experiments have evaluated each nutrient's contribution to an overall model accuracy by simulating the removal of each nutrient one at a time from the input feature vector. Removing nitrogen resulted in a decrease in accuracy from 99.09% to 91.3%, and removing only potassium reduced accuracy to approximately 86%. Therefore, the contribution of NPK levels as the strongest single factor influencing model predictions is in agreement with agronomists' preconceived notion that availability of macronutrients is the most important factor in determining whether or not a crop will grow in any specific soil profile.

Module 2 (geospatial prediction) resolves 96.4% of agricultural districts in India (246,091 training records) for district/ season/ crop queries through state/ season majority class (for 3.6%). This is better than any district-level advisory system reported in the literature. Joblib serialisation saves <100ms response time for each query on low-specification servers by avoiding complete reconstructions of the full tree structure on each request.

Module 5 (profit analytics) closes the decision loop by grounding agronomic recommendations in financial reality. For all recommended crops, the system provides season averages for yield, current market prices, benchmark costs of cultivation, and projected net margins per hectare at both national and the top 8 states' level. None of the systems surveyed in the literature (including Kumar and Kumar's [2] very high accuracy framework) provided this financial framing, but it could perhaps be considered the most important output when subsistence farmers are considering what to plant.

## IX. CONCLUSION

This document described the introduction of the Integrated Cropping Management System, a spatial agriculture platform that gives farmers access to five crop advisory modules based on machine learning — Random Forest for soil matched crop recommendation (99.09% accuracy); Gini Decision Tree for geospatial crop prediction (246,091 district records for 99.0% accuracy); label encoded Decision Tree from scikit-learn for fertilizer prescriptions need (92.6% cross validated accuracy); monthly rain prediction using the mean for the past 115 years (mean absolute error = 18.4 mm/month); and yield/thought analysis for financial record keeping using pandas GroupBy aggregation. All of these were implemented into one web-based PHP application that can be installed on a commodity server. A PHP/Python architecture enables the system to be implemented without using GPUs (they are not required), without cloud infrastructure (this is not a requirement) and without the need for relational database systems; therefore, the integrated crop management system can be installed by agricultural cooperatives/offices that lack the resources to install and maintain standalone IT systems.

The ICMS provides a solution to a structural gap present in the current literature. While numerous existing systems, such as the high-performing, highly accurate, XAI-enabled illustration provided by Kumar and Kumar [2]; contribute to the ability of an agronomist to make an accurate agronomic decision by providing information about the agronomic, geographical, nutritional, climatic, and financial appraisal of any one of five Agronomic decisions prior to planting. None of the systems in existence to date provide information on all five of these appraisal dimensions at the same time (as a decision-support system). Thus, the profit analytics layer is a means for users to extend the functionality of the system beyond simply agronomic optimization into the area of business decision-making. Specifically, it enables growers to evaluate season-to-season yield projections (state-by-state) and evaluate the expected costs of growing their crops (state-by-state) and evaluate the anticipated market prices for those crops (state-by-state) with regard to the ultimate projected net income to be derived from the establishment of their crops.

Future development can be made by: (i) Incorporating IoT Field Sensors for Real-time Measurement of NPK and Soil Moisture Levels rather than allowing users to enter static values; (ii) Using the LIME and SHAP XAI Framework of Kumar and Kumar [2] to Provide Farmers with Explanations Based on Feature Level for Each Recommendation; (iii) Migrating the Python Inference Layer to a Flask or FastAPI Web Service to Support More Concurrently Connected Users; (iv) Ask Farmers to Help Generate the 99-record Fertiliser Dataset which Currently Limitations on the Amount of Data Available for the System; (v) Including NDVI and EVI Vegetation Indices from Satellites Servicing to Augment the Point Estimate of Yield with Spatially Distributed Information; and (vi) Develop a Progressive Web Application Shell with Embedded TensorFlow.js to Assist Producers in Areas of Intermittent Internet Service.

## REFERENCES

[1] Report of annual agriculture and farmer's welfare from Indian government on agricultural statistics from Directorate of Economics and Statistics dated November 2023.
[2] S. Kumar & M. Kumar, An AI (Explainable AI) based agriculture advice program based off of previous crop yields by soil conditions to provide farmers accurate information on planting and crop selection using IEEE publication, Volume 71 Issue 2, pages 6950-6959 dated May 2025, doi:10.1109/TCE.2025.3569736
[3] O. Turgut, I. Kok, & S. Ozdemir (2024), The architecture integrated (AgroXAI) uses AI based algorithms to create agriculture advisory systems. Proceedings of 2024 IEEE Int. Conference on Big Data (BigData), pages 7208-7217, doi:10.1109/BigData62323.2024.10825641
[4] M.Y. Shams, S.A. Gamel, & F.M. Talat (2024), Our research builds upon the previous work of others to propose the use of Explainable Artificial Intelligence (EAI) within decision-making for agriculture. In Neural Comput. Applications Volume 36, Issue 11, pages 5695-5714 & doi:10.1007/s00521-023-09391-2
[5] A. Khan & M. Faheem, R.N.Bashir, C. Wechtaisong, & M.Z. Abbas (2023), Artificial Intelligence and Internet of Things (IoT) technologies support sustainable agriculture and smart farming. IEEE Access Volume 11 Pages 78686-78692, doi:10.1109/ACCESS.2023.3298215
[6] D. Elavarasan & P.M. D. Vincent, Deep Reinforcement Learning based model for Crop Yield Prediction for Sustainable Agronomical Applications. IEEE Access Volume 8 Pages 86886-86901, doi:10.1109/ACCESS.2020.2992480
[7] A.Khan, M.Faheem, R.N.Bashir, C.Wechtaisong & M.Z.Abbas (2022), Internet of Things (IoT) based on contextual awareness offers fertiliser recommendations. IEEE Access Volume 10 Pages 129505-129519 doi:10.1109/ACCESS.2022.322816
[8] V. Kale and B. N. Mohapatra, "Crop Recommendation System Using Machine Learning," ITEGAM - JETIA, vol.10, no.48, pp 63-68, Jul/Aug 2024, doi:10.5935/jetia.v10i48.1186.
[9] D. Gosai, C. Raval, R. Nayak, H. Jayswal and A. Patel, "Crop Recommendation System Using Machine Learning," International Journal of Scientific Research in Computer Science and Engineering, vol.7, no.3, pp 554-569, May/Jun 2021, doi:10.32628/CSEIT2173129.
[10] M. Bouni, B. Hssina, K. Douzi and S. Douzi, "Integrated IoT Approaches for Crop Recommendation and Yield Prediction Using Machine Learning," IoT, vol.5, no.4, pp 634-649, 2024, doi:10.3390/iot5040028.
[11] S. Shastri, S. Kumar, V. Mansotra et Al., "Advancing Crop Recommendation System with Supervised Machine Learning and Explainable Artificial Intelligence," Scientific Reports, vol.15, article 25498, 2025, doi:10.1038/s41598-025-07003-8.
[12] Y. Akkem, S. K. Biswas and A. Varanasi, "Role of Explainable AI in Crop Recommendation Technique of Smart Farming," International Journal of Intelligent Systems and Applications, vol.17, no.1, pp 31-52, Feb 2025, doi:10.5815/ijisa.2025.01.03.
[13] T. Mahesh and R. Soundrapandiyan, "Yield Prediction for Crops by Gradient Based Algorithms," PLOS ONE, vol 19, no.8, e0291928, Aug 2024, doi:10.1371/journal.pone.0291928.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[14] M. K. Senapaty, A. Ray and N. Padhy, "A Decision Support System for crop recommendation using machine learning classification algorithms," Agriculture, vol.14, no.8, article 1256, 2024 doi:10.3390/agriculture14081256
[15] The article "Incorporating soil information with machine learning for crop recommendation to improve agricultural output" was written by H. Afzal, M. Amjad, A. Raza, and others. It was published in the scientific journal Sci. Rep. on March 2025. It can be accessed using the DOI number 10.1038/s41598-025-88676-z.
[16] The second article by R. A. Ahmed, W. El-Shafai, Z. A. Ahmed, E. M. El-Rabaie, and F. E. Abd El-Samie is titled "High-Precision Crop Recommendation System With Stacking Ensemble Classifiers for Optimizing Agricultural Productivity" and was published in December 2025 in the same journal as above. Access can be gotten through DOI number 10.1038/s41598-025-09640-5
[17] The third article "Machine Learning based Recommendation of Agricultural & Horticultural Crop Farming in India under the Regime of NPK, Soil pH and 3 Climatic Variables" was written by B. Dey, J. Ferdous, and R. Ahmed. It appeared in Heliyon on February 2024. Access is available using DOI number 10.1016/j.heliyon.2024.e25112.
[18]"Random Forests" by L. Breiman, published in Machine Learning (Vol. 45, No. 1, pp 5-32) in October of 2001. (ISSN 0885-6125) DOI 10.1023/A:1010933404324
[19] A Mathematical Theory of Communication by C. E. Shannon , published in Bell System Tech Journal (Volume 27, No. 3) July, 1948 (ISSN: 0005-8580) doi:10.1002/j.1538-7305.1948.tb01338.x
[20]"Scikit-learn: Machine Learning in Python" by F. Pedregosa and others, published in J.M.L.R. (v. 12 pp 2825-2830), 2011. Access via https://jmlr.org/papers/v12/pedregosa11a.html.
[21] A. Ingle, “Crop recommendation dataset,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset
[22] Abhinand, “Crop production in India,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/abhinand05/crop-production-in-india
[24] G. D. Abhishek, “Fertilizer prediction,” Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction
[25] Rajanand, “Rainfall in India — sub-division wise monthly data 1901–2015,” Kaggle, 2017. [Online]. Available: https://www.kaggle.com/datasets/rajanand/rainfall-in-india

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING