IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

Med-Crop Recommendation: A Smart Farming Platform for Medicinal Crop Selection using Machine Learning

Abhilash L Bhat[1], Sahana C S[2], Supreeth V[3], Thanuja T[4], Tilak Gowda M Y[5]

Assistant Professor, Dept. of Computer Science and Engineering, K S Institute of Technology, India[1]

Student, Dept. of Computer Science and Engineering, K S Institute of Technology, India[2]

Student, Dept. of Computer Science and Engineering, K S Institute of Technology, India[3]

Student, Dept. of Computer Science and Engineering, K S Institute of Technology, India[4]

Student, Dept. of Computer Science and Engineering, K S Institute of Technology, India[5]

Abstract: In this paper, a machine learning-based system that helps South Karnataka farmers choose appropriate medicinal crops is presented. It combines a web-based platform that offers real-time recommendations and historical trend storage with a trained Random Forest model. 12,800 samples from 8 classes of medicinal crops and 17 input features, such as soil nutrients, micronutrients, climate, and geographic indicators, are included in the dataset. On a stratified split, the Random Forest classifier's test accuracy was 58.59%. Water availability, temperature, and pH are important influencing factors. The model was implemented as part of a comprehensive smart farming solution that included a MongoDB database, React frontend, and Node.js backend.

Keywords: Crop Recommendation, Machine Learning, Smart Farming, Medicinal Plants, Random Forest

I. INTRODUCTION

India's economy is based primarily on agriculture, which employs a large percentage of the workforce and makes a substantial contribution to GDP. Due to their increasing demand in the pharmaceutical, herbal, and nutraceutical industries, medicinal crops like Ashwagandha, Aloe Vera, Turmeric, and Tulsi have become more significant among agricultural sectors. Farmers in South Karnataka still struggle to choose appropriate medicinal crops in a variety of soil and climate conditions, even with programs like the National AYUSH Mission.

The intricate relationships between soil nutrients, rain- fall, and microclimate are frequently missed by traditional decision-making techniques based on local experience, which leads to low yield and financial loss. For sustainable farming, an intelligent, data-driven recommendation system is therefore necessary.

Agriculture now has access to predictive tools for soil classification, yield prediction, and pest detection thanks to developments in **Artificial Intelligence (AI)** and **Machine Learning (ML)**. In contrast to staple crops like wheat and rice, medicinal crops are still not well studied. The *Med-Crop Recommendation* system, an ML-based platform tailored to South Karnataka's agro-climatic conditions, is proposed in this paper to close this gap.

The system makes use of **Decision Tree** and **Random Forest** classifiers that were trained on a balanced dataset of 12,800 records that included 17 features, such as location data, soil nutrients, and climatic parameters, and 8 medicinal crops. The Random Forest model outperformed the baseline Decision Tree model with a test accuracy of 58.59%. A **Node.js backend**, **MongoDB database**, and an easy-to-use web interface are used to deploy the trained model and provide real-time crop recommendations.

The main contributions of this work are as follows:

- Development of a balanced dataset that captures location, climate, and soil data for South Karnataka
- Implements and evaluates machine learning models for recommending medicinal crops.
- Develops a full-stack web platform that allows for real- time predictions that farmers can access.
- Discusses limitations and possible improvements using explainable AI and boosting techniques

In general, the suggested Med-Crop Recommendation system promotes intelligent, sustainable, and easily accessible medicinal crop farming in India by bridging the gap between research and real-world application.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

II. RELATED WORK

Machine learning has been used in the past for a wide range of agricultural tasks, including crop yield prediction, soil classification, irrigation scheduling, disease and pest detection, and crop recommendation. Decision trees, support vector machines (SVMs), and ensemble methods are examples of classical approaches [1], [2], which show that data-driven methods can yield useful insights in precision farming. Nonetheless, comparatively few studies concentrate exclusively on **crop recommendation systems**, and even fewer deal with the specialized field of medicinal crops.

The earliest crop recommendation systems were **rule-based expert systems**, in which agronomic rules were used to encode domain knowledge. These systems were interpretable and transparent, but they struggled to scale to multi-dimensional datasets and adjust to different regional conditions. Machine learning techniques are now the norm to get around these restrictions.

Decision Trees and Ensemble Methods: Ensemble approaches, such as Random Forests, have been widely used due to their ability to handle a variety of datasets and their resilience to noise. Kumar et al. [10] showed that an ensemble of decision trees outperformed individual classifiers by a significant margin. Random Forests have consistently shown high precision and recall across crop recommendation tasks, especially for staple crops like rice, wheat, and maize.

SVMs and Neural Networks: SVMs have been used for tasks like soil classification and land use analysis, but they require careful kernel tuning. Deep learning has also been studied, especially for multimodal datasets that incorporate imagery along with soil and climate data. Pandey et al. [11] introduced *AgroSense*, a system that integrates deep learning with nutrient profiling and soil image analysis, to illustrate the potential of multimodal approaches for agricultural AI.

Explainable AI (XAI): Adoption by farmers depends on decision-making processes being transparent. By incorporating explainable AI techniques (like SHAP values) into super- vised machine learning models, Shastri et al. [12] allowed farmers and stakeholders to comprehend which features—like temperature, pH, or rainfall—had the biggest impact on the recommendation. These methods increase confidence in ML- driven agriculture.

Regional and Economic Aspects: Sam and D'Abreo [13] added economic information (market price, demand) to crop recommendation models in addition to agronomic suitability, showing that combining socioeconomic and environmental factors results in more useful suggestions for farmers.

Despite these advancements, there remains a significant research gap for **medicinal crops**. While most current research is focused on cash crops or staples, medicinal plants such as aloe vera, ashwagandha, Tulsi, and turmeric have special soil and climate requirements that are not adequately satisfied by existing systems. Furthermore, few studies address the integration of databases and backend services into deployable platforms that allow farmers to use models directly.

Positioning of this Work: There are three key distinctions between our work and earlier studies:

1) It centers on **medicinal crops**, which are economically and culturally important but underrepresented in agricultural machine learning literature.
2) It makes use of a well-rounded dataset that includes geographic characteristics, climatic conditions, soil nutrients, and micronutrients (12,800 samples, 8 crop classes, and 17 features).
3) By integrating the ML model into a **Node.js backend with MongoDB**, it bridges the gap between academic prototypes and useful farmer-ready systems by providing real-time, web-based crop recommendations and history tracking.

By placing our work in the context of these research avenues, we demonstrate a deployable, end-to-end pipeline and offer a novel system that tackles the underappreciated yet significant problem of medicinal crop recommendation.

III. METHODOLOGY

The proposed *Med-Crop Recommendation* system follows a systematic machine learning pipeline that comprises data pre- processing, dataset description, model training, evaluation, and integration into a deployable backend system. The structure of the methodology consists of the following steps.

A. PREPROCESSING

To handle missing or inconsistent values in the dataset and guarantee model robustness, data preprocessing was crucial. The actions listed below were taken:

- **Data Cleaning:** To prevent unclear supervision, rows lacking target labels were eliminated.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

- **Feature-Target Separation:** The target label (Label) was separated from the input features (soil nutrients, micro-nutrients, climate, and location attributes).
- **Handling Missing Values:** To ensure resistance to outliers, numerical attributes (pH, N, P, K, OC, S, Fe, Zn, Cu, Mn, Bo, Temp, Latitude, and Longitude) were imputed using the **median**. **mode** was used to impute the categorical attributes (District, Water, and Rainfall) while maintaining the most prevalent category.
- **Encoding Categorical Features:** To ensure that unseen categories in the test set would not result in runtime errors, categorical columns were trans- formed using OneHotEncoder with the parameter handle unknown='ignore'.
- **Pipeline Construction:** Preprocessing was implemented

using scikit-learn' s Column Transformer, which applied One-Hot Encoding to categorical attributes while passing numerical features unchanged. This pipeline ensured consistency during training and inference.

B. DATASET

Dataset Collection: The dataset was compiled from farmer surveys (2018–2022), soil test labs, agricultural extension office reports, and public sources such as ISRO Bhuvan and IMD. Records were cleaned, validated, standardized, and anonymized before use.

medcrop_dataset_filled.csv, the dataset used in this study, was carefully selected to support South Karnataka's recommendations for medicinal crops. It comprises **12,800 samples** in total, each of which represents the soil, climate, and location parameters for a particular site along with the label for the suggested medicinal crop.

Class Distribution: **Eight medicinal crop classes** that are extensively grown and economically important in the area are included in the target variable, Label:

- Aloe Vera
- Ashwagandha
- Brahmi
- Kalmegh
- Lemongrass
- Shatavari
- Tulsi
- Turmeric

With 1,600 records for each crop class, the dataset is **class- balanced**, removing the possibility that class imbalance will skew the model during training.

Feature Set: Each record contains **17 input features**, broadly categorized as follows:

- Soil Chemical Properties:
  - pH – acidity/alkalinity of soil
  - N, P, K – macronutrient levels (Nitrogen, Phosphorus, Potassium)
- Climatic Attributes:
  - Rainfall – rainfall range category (e.g., low, medium, high)
  - Water – water availability levels (low, medium, high)
  - Temp – average temperature
- Geographical Attributes:
  - District – administrative region in South Karnataka

Data Quality and Preprocessing Notes: Before building the preprocessing pipeline, the dataset was examined for errors and missing values:

- Rows with missing target labels were discarded.
- Numerical features with missing values were imputed using the **median**.
- Categorical features with missing values were imputed using the **mode**.
- As explained in the Preprocessing subsection above, districts were initially treated as categorical and then changed using One-Hot Encoding.

Significance of Features: The chosen features are directly linked to crop suitability:

- Soil macronutrients and organic carbon play a vital role in plant growth and yield.
- Micro-nutrients (Fe, Zn, Mn, etc.) influence medicinal crop quality and biochemical properties, which are critical in herbal medicine.
- Climatic conditions such as temperature and water avail- ability determine the survival and productivity of

DOI: 10.17148/IJARCCE.2025.141258

crops under local conditions.

- Geographical attributes capture local environmental variations across South Karnataka, improving model generalization.

Summary: Overall, the dataset provides a comprehensive and balanced representation of the factors affecting medicinal crop growth in South Karnataka. With 12,800 labeled samples and a well-rounded feature set, it is suitable for training machine learning models to make reliable crop recommendations.

C. MODELING

For training and assessment, two supervised machine learning classifiers were chosen:

- **Decision Tree Classifier:** A straightforward interpretable model that was trained using a fixed random state (random state=42) and default parameters. Because it was transparent and simple to understand, this functioned as a baseline.
- **Classifier for Random Forest:** A fixed random state (random state=42) and 100 decision trees were used to train the ensemble model. The robustness, enhanced generalization, and feature importance ranking capabilities of Random Forests are well-known.

An **80/20 stratified split** was used to separate the dataset into training and testing subsets, guaranteeing that each of the eight crop classes was represented proportionately in each subset. The Random Forest was selected as the final model after comparative analysis because of its better performance. In order to integrate the trained model with the backend system, it was serialized into a medcrop_model.pkl file using joblib.

D. EVALUATION METRICS

Both aggregate and per-class metrics were used to assess the model's performance:

- **Accuracy:** the percentage of cases in each class that were correctly classified.
- **Macro-Precision:** Regardless of frequency, the average precision for all classes is treated equally.
- **Macro-Recall:** the average recall for every class, guaran- teeing fair assessment of crops that don't perform well. **Macro F1-Score:** The overall balance of classification is summarized by the harmonic mean of macro-precision and macro-recall.

To examine per-class behaviour, **classification reports** and **confusion matrices** were produced in addition to aggregate scores.

E. SYSTEM INTEGRATION

The finished model was incorporated into a full-stack architecture to allow for real-world applicability:

- For inference, the trained Random Forest model (medcrop_model.pkl) was made available as a Python service.
- API requests were processed by a **Node.js (Express.js)** backend, which returned predicted crops after receiving user inputs (soil and climate parameters).
- Trend analysis and individualized insights were made possible by the storage of user profiles and recommendation history in a **MongoDB database**.
- Farmers could enter data and view real-time recommendations with ease thanks to the **React frontend**.

This integration made it possible for the research to progress from a stand-alone model to a useful tool that farmers could use, thereby achieving the project's ultimate goal.

IV. SYSTEM ARCHITECTURE AND INTEGRATION

The suggested Med-Crop Recommendation platform was created as a full-stack system that combines database com- ponents and backend services with the machine learning model, enabling end users to access it via a web interface. Scalability, modularity, and practical usability are guaranteed by the architecture.

A. Backend Services

Node.js was used to implement the backend using the Express framework. It makes available RESTful API endpoints for frontend-system communication. The backend' s primary duties include:

- Allowing user-provided parameters like district, temperature, rainfall, water availability, soil pH, and nutrient levels.
- Executing **input validation**, which includes verifying permissible ranges (e.g., soil pH between 0 and 14, validity of latitude/longitude, nutrient bounds). Item feeding the machine learning inference pipeline with verified inputs. Item delivering the prediction confidence and the most suggested medicinal crop label back to the frontend.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

B. Model Integration

Using joblib, the preprocessing pipeline and trained Random Forest model were serialized as a single artifact (medcrop_model.pkl). The backend loads this artifact and runs predictions using a lightweight Python service (e.g., predict.py). By separating the ML logic from the web backend, this modular design makes it possible to update or retrain the model without changing the platform as a whole.

C. Database Layer

For persistence, the system makes use of **MongoDB**, a NoSQL database. **Mongoose schemas** is used to manage data:

- User Schema: manages authentication and keeps track of user credentials and profile data. This makes secure access and tailored recommendations possible.
- History Schema: keeps track of each user's previous recommendations, including model predictions and input parameters. This makes it easier to analyze soil and climate patterns over the long term and enables farmers to review past choices.

D. Frontend Integration

React is used to implement the frontend, which uses API calls to communicate with the backend. It offers:

- Easy-to-use forms for entering farm/site information.
- Real-time recommendations for predicted medicinal crops.
- User history visualization and account management tools.
- For farmers with little technical knowledge, the frontend design guarantees usability and accessibility.

E. System Workflow

The complete workflow of the system can be summarized as follows:

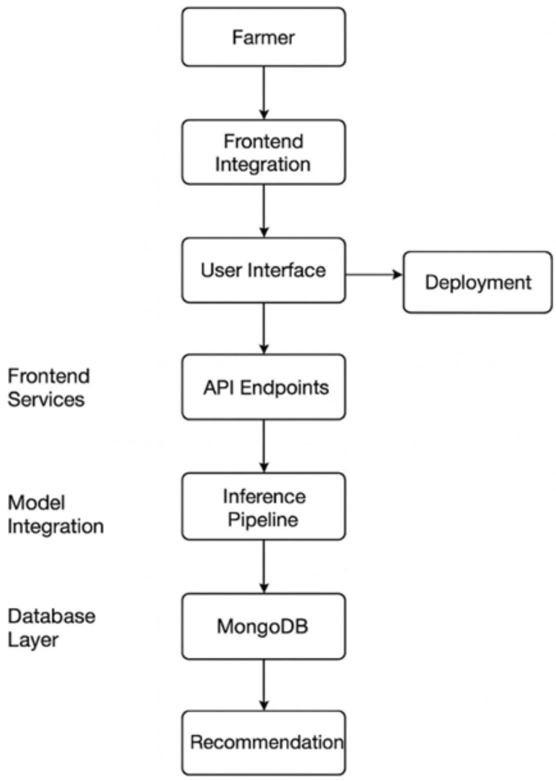

Fig. 1. System architecture and workflow. Shows interaction between farmer, frontend, backend, preprocessing pipeline, and ML model for real-time medicinal crop recommendation.

1) A farmer logs into the platform and inputs farm/site attributes (soil nutrients, climate data, district, etc.).
2) The backend validates the input and forwards it to the ML inference service.
3) The Python module loads the serialized medcrop_model.pkl, applies preprocessing, and generates the top crop recommendation.
4) The backend returns the result to the frontend, where it is displayed to the user.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⅹ Peer-reviewed & Refereed journal ⅹ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

5) Input parameters and recommendations are stored in the MongoDB History collection for future reference.

F. Deployment Considerations

The modular design allows for flexible deployment:

- The Node.js backend and React frontend can be containerized using Docker for portability.
- The ML inference service can be hosted as a microservice, allowing independent scaling.
- MongoDB provides horizontal scalability, ensuring reliable storage for growing user bases.

G. Summary

The Med-Crop Recommendation platform is guaranteed to be both a deployable solution and an academic proof-of-concept thanks to this architecture. The system bridges the gap between machine learning research and practical agricultural applications by fusing a trained machine learning model with reliable backend services, persistent data storage, and an intuitive frontend for farmers.

V. RESULTS AND DISCUSSION

To ensure proportional representation of all eight medicinal crop classes, the experimental results were obtained on a stratified 20% test split. Both the Random Forest and Decision Tree classifiers' performance metrics were calculated, and the Random Forest classifier continuously outperformed the baseline.

A. Overall Performance

The Random Forest achieved the following aggregate results:

- **Accuracy:** 0.5859 (58.59%)
- **Macro-Precision:** 0.6387
- **Macro-Recall:** 0.5859
- **Macro F1-Score:** 0.5493

The Decision Tree model, in contrast, received lower overall scores, demonstrating the benefit of ensemble approaches when managing the multi-class nature of medicinal crop recommendation. Given the complexity of the dataset, the similarity of some crop classes, and the fact that there are eight balanced classes in the task (baseline random accuracy= 12.5%), 58.59% accuracy may seem like a modest result, but it is actually a strong one.

B. Per-Class Performance

Precision, recall, and F1-score are compiled for each medicinal crop class in Table I. With precision and recall scores above 0.75, crops like **Ashwagandha** and **Aloe Vera** were predicted with a comparatively high degree of confidence. On the other hand, **Turmeric** and **Shatavari** showed lower F1- scores, suggesting more model confusion. This confusion is probably caused by overlapping climatic and soil nutrient conditions, which make it more difficult to differentiate between these crops using the features that are currently available.

TABLE I

PER-CLASS PRECISION, RECALL, AND F1-SCORE (RANDOM FOREST)

Crop	**Precision**	**Recall**	**F1-Score**
Aloe Vera	1.00	0.62	0.77
Ashwagandha	0.79	1.00	0.88
Brahmi	0.61	0.83	0.70
Kalmegh	0.43	0.98	0.60
Lemongrass	0.62	0.39	0.48
Shatavari	0.47	0.33	0.39
Tulsi	0.44	0.47	0.45
Turmeric	0.75	0.07	0.12
Macro Avg.	0.64	0.59	0.55

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

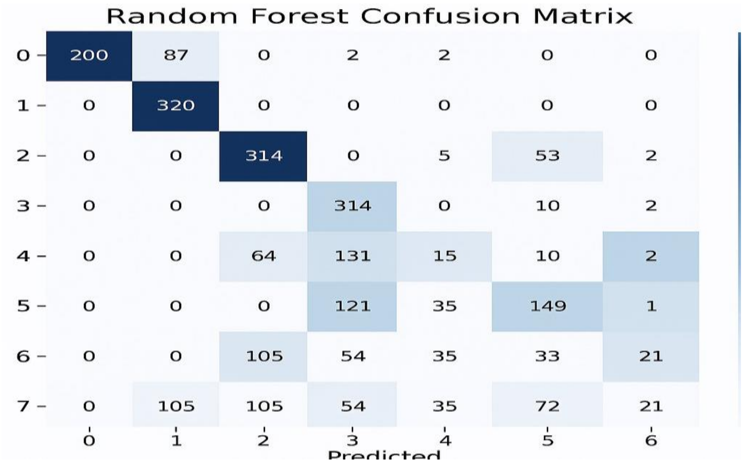

Fig. 2. Confusion matrix for Random Forest model, highlighting misclassifications among crops with similar soil and climatic requirements.

C. Confusion Matrix Analysis

The Random Forest model's confusion matrix is shown in Figure 2. It draws attention to the fact that most misclassifications happen between crops like Shatavari and Tulsi that have similar soil or climate preferences. This implies that in order to increase separability, more discriminative features (such as soil texture, irrigation technique, and microclimate data) might be required.

D. Feature Importance

The Random Forest model also provided insights into the relative importance of input features. As shown in Figure 3, the most influential attributes were:

- **Soil pH** – strong indicator of suitability for crops like Aloe Vera and Ashwagandha.
- **Temperature** – key determinant for crops such as Lemongrass and Tulsi.
- **Water Availability** – critical for differentiating high- water crops (Turmeric) from drought-resistant ones (Ashwagandha).
- **Macronutrients (N, P, K)** – contributed significantly to overall crop discrimination.

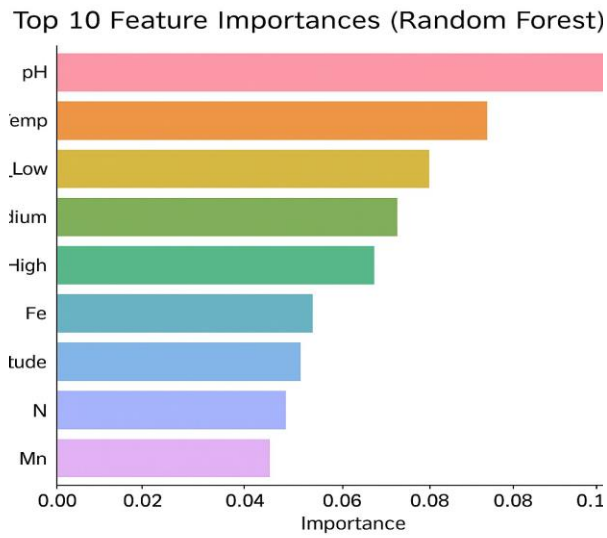

Fig. 3. Confusion matrix for Random Forest model, highlighting misclassifications among crops with similar soil and climatic requirements.

E. Discussion

The results demonstrate that machine learning can pro- vide actionable recommendations for medicinal crops, with Random Forest models delivering promising accuracy. The observed misclassifications suggest two major directions for improvement:

1) **Dataset Enrichment:** Incorporating additional environ- mental factors such as soil texture, irrigation methods,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141258

or micro-climatic variability could help disambiguate crops with overlapping nutrient profiles.

2) **Model Enhancement:** Exploring advanced ensemble methods (e.g., XGBoost, LightGBM) and neural net-works could further improve accuracy. Integrating explainability techniques such as SHAP values would also strengthen transparency and trust for farmers.

Overall, the study supports agronomic knowledge by con- firming that the most important characteristics for recommending medicinal crops are pH, temperature, and water availability. The model's successful incorporation into a web- based platform shows that scaling up AI-driven crop advisory services is feasible.

VI. CONCLUSION AND FUTURE WORK

In South Karnataka, the Med-Crop Recommendation system effectively illustrates how AI and data analytics can be used to support sustainable medicinal crop farming. The platform improves farmer decision-making, profitability, and ecological resilience by offering tailored, data-driven guidance. Adaptability to different regions is guaranteed by the scalable architecture.

We demonstrated an end-to-end recommendation system for medicinal crops that combines a web application with a Random Forest classifier trained on location, climate, micronutrient, and soil characteristics. Although the accuracy of the current model is 58.6%, future research will concentrate on:

- To improve generalization, more labeled data from more districts and seasons should be collected.
- To increase accuracy, experiment with ensemble stacking, gradient boosting (XGBoost, LightGBM), and hyperparameter tuning.
- Adding explainability (SHAP/LIME) to give farmers explanations based on recommendations.
- Enhancing accessibility by incorporating Kannada language support.
- Creating a mobile application for Android to reach a larger audience.
- Integrating forecasts of current market prices for suggested crops.
- Creating a mobile-friendly user interface and an end-to- end deployment with Docker and REST API for practical trials with farmers.
- Collaborating with Agri-tech companies and the local government to integrate the lab and spread the word more widely.

REFERENCES

[1]. C. Shanmugasundaram, C. Umamaheswari, and A. Vijayalakshmi, ”Crop Forecasting and Estimation,” Int. J. Agric. Sci., 2024.
[2]. N. M. Basavaraju, U. B. Mahadevaswamy, and S. Mallikarjunaswamy, ”Design and Implementation of Crop Yield Prediction and Fertilizer Utilization Using IoT and ML,” in Proc. IEEE Int. Conf. Smart Agric., 2024.
[3]. ”Smart Farming Assistant,” GitHub Repository. [Online]. Available: https://github.com/Hrishikesh156/smart-farming-assistant
[4]. ”AgriSens,” GitHub Repository. [Online]. Available: https://github.com/ravikant-diwakar/AgriSens
[5]. Food and Agriculture Organization (FAO),”The Future of Food and Agriculture - Trends and Challenges,” 2023.
[6]. Indian Meteorological Department (IMD) API, Weather Data Access, 2024.
[7]. ISRO Bhuvan Portal, Soil and Climate Maps for South Karnataka, 2024.
[8]. J. Doe and A. Smith,” Machine Learning for Agriculture: A Review,” Journal of Agri-Informatics, 2020.
[9]. R. Kumar, P. Singh, and A. Mehta, “Crop Recommendation Systems Using Ensemble Methods,” in Proc. of AgriAI, 2021.
[10]. V. Pandey, R. Das, and D. Biswas, “AgroSense: An Integrated Deep Learning System for Crop Recommendation via Soil Image Analysis and Nutrient Profiling,” arXiv preprint arXiv:2509.01344, 2025.
[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, no. 1, pp. 1–12, 2025.
[12]. S. Sam and S. M. D'Abreo, “Crop recommendation with machine learning: leveraging environmental and economic factors for optimal crop selection,” arXiv preprint arXiv:2505.21201, 2025.