Grenze International Journal of Engineering and Technology, June Issue

Crop Recommendation Model using Machine Learning based on Soil Analysis

Lata Tembhare[1], Harsh Ladwani[2], Anurag Bansod[3], Himanshu Urkude[4], Mayur Hedau[5] and Naman Gore[6]
[1-6]Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India
Email: latabopche111@gmail.com, harshladwani18@gmail.com, anuragpbansod@gmail.com, himanshuurkude21@gmail.com, mayurhedau57@gmail.com, namangore161@gmail.com

***Abstract*— Agriculture is central to international food security but it is coming under increasing pressure from climatic change, soil erosion and inefficient use of inputs. This study introduces a machine learning based crop recommendation mechanism that replaces experience-based crop selection with decision support mechanism based on data provided by key soil and weather features that include Nitrogen (N), Phosphorus (P), Potassium (K), pH, rainfall, humidity and temperature. Multiple supervised learning models are combined using stacked ensemble strategy to improve the prediction stability especially the closely related crop classes. A modular and connectivity aware input layer provides for the use of inputs for both manual data entry, and sensor based data acquisition, and for use in environments with limited or with intermittent internet connectivity. Experimental evaluation on publicly available Kaggle Crop Recommendation data set has shown as Random Forest classifier the accuracy of test dataset reaches 99.55% and ensemble model achieves 99.32% in lower variance validation coefficient which shows improvement in model robustness. Results proving that the suggested framework is effective and reproducible and well adapted to precision agriculture and to future IoT-enabled smart farming systems.**

***Index Terms*— Crop recommendation, ensemble learning, precision agriculture, soil analysis, machine learning and smart farming.**

I. INTRODUCTION

Nearly 45% of India's workforce rely on agriculture as a source of livelihood therefore, this is an important sector with some 17.7% of Gross Value Added (GVA), of country being given by this sector in 2023-24. However, the sector faces severe challenges including soil erosion, water table depletion and on increasing variability of climatic conditions. These pressures are leading to yield losses, inconsistency of quality and economic uncertainty to farmers. A major driving force is the fact that overall selection of crops is based oftentimes on personal experience or tradition and not supported by data to achieve the best outcome with a minimum use of inputs. This work fills the gap by a next generation, machine learning crop recommendation framework, with the focus of precision, interpretability and robustness. In contrast to the systems based on one single classifier use multiple algorithms are used and combined (in a way, learned using weights - stacked ensemble) for soil and weather features. This ensemble help to improving the stability between closely related crop classes and weaken the single model weakness to get good and high accuracy prediction.

Grenze ID: 01.GIJET.12.2.483_22

The framework runs on important features of the soil and meteorology, like N, P, K, pH, rainfall, humidity and temperature to make context-aware, evidence-based recommendations. The system architecture comprises an input layer which is connection aware and modular, to enable manual data entry devices as well as plug-in sensor devices which enable the system to operate on slow or sporadic networks. It provides operational framing well - a single, benchmarked pipeline where the public datasets validation exists though it is more focused on digesting information of the live sensors to enhance reproducibility and readiness to deploy. As the existing model directly relates to the accuracy of the supervised learning in terms of recommendations, its design is currently modular, thus it would not complicate extending the technology with IoT devices or deep learning, in addition to remote-sensing data. The computational background of this work paves the way on the development and application of smart agriculture systems in the future that would be able to help with the near realistic decision making and adaptive management of smart agriculture systems. The remainder of the paper is divided in the following manner: Section II provides literature review, Section III outlines the proposed solution, Section IV provides methodology, Section V provides results and discussion, Section VI provides conclusion and finally, Section VII provides the future work.

II. Literature Review

The existing developments in precision farming have indicated that the domains of the Machine Learning (ML) and Internet of Things (IoT) technologies are highly effective in enhancing crop output and supporting sustainable agriculture. Numerous studies have been done on the smart systems that direct farmers based on intelligent systems taking into account the characteristics of soil, weather conditions, and sensor data.

Other researchers have focused on the IoT-based system that uses ML techniques to detect soil conditions and give recommendations on crops. They are generally sensor networks that can be utilized in measuring the moisture of soil, its temperature, and the amount of the nutrients in the soil so that analysis and decision making can occur automatically [1][2].

Although these methods emphasize on real-time monitoring of the soils, most are constrained by the fact that they fail to take into consideration the dynamic weather contents and are mostly experimented in controlled or simulated environments.

The predictive modeling of rainfall, crop yield and soil nutrient estimation based on machine learning and deep learning methods has been investigated in other studies [3], [4], [8]. Such publications have high prediction capabilities, but they are aimed primarily at prediction rather than at giving the farmers practical implications of what they are to grow regarding farming. Moreover, deciphering and adjusting and adapting of intricate models to other agricultural areas is an issue.

Artificial intelligence systems have also been suggested as platforms to participate in crop advisory and soil analysis systems [6] [7] using deep learning and massive labeled data. These techniques have a high degree of accuracy, but require large quantities of annotated data, rendering their useful usage to regions where the database is available, or hard to access.

Several studies have explored the integration of multiple machine learning models into a single system to form a agricultural decision support [5], and also research has been done on feature selection strategies and evolutionary or generative models of soil and climate modeling [9] [11].

In spite of such developments, majority of the existing systems are used in offline settings or do not have smooth operations of integrating diverse in nature of data and thus their efficiency is not achieved in real time situations.

Cloud computing, land use, geospatial, and neural network-based image models such as CNNs and RNNs have also been introduced as recent studies in the image domain and large-scale agricultural utilization cases [13] [18]. Nevertheless, real-time data on soil and live data on weather, as well as adaptive recommendation systems and operationalizing them into a scalable system, are still limited.

On this basis, there is necessity of crop recommendation system that can take into account both soil and weather data in real time, is simple to update and give consistent and interpretable recommendations. The proposed work fills this gap with an all-in-one system of multi-source data integration and machine learning grounded decision support in enhancing effective applications in various farming environments.

III. Proposed Approach

This strategy is used to promote decisions to farmers, on which kind of crops to cultivate by integrating data on soil (N, P, K, pH) along with weather data that are local to the farmland as well as interpreting that transform evidence into easy and practical recommendations. Besides giving recommendations on crop choice, the system provides viable and succinct advice like sowing windows and irrigation or fertilizer tips to act on data in order to make a decision.

A. System overview

Data Acquisition Layer
This layer acquires inputs related to soil and weather from either sensors and/or basic forms- Nitrogen (N), Phosphorus (P), Potassium (K), pH, rainfall (mm), humidity (%) and temperature (°C).

Data Processing and Validation Layer
This layer validates the quality of the data, interpolates missing values, eliminates outliers, scales all variables down to a common scale, and then modifies the class balance so that the models do not have a bias towards the most common cropping classes. Later, the data is used in the training and prediction layers.

Prediction Layer
This layer utilize a range of classifiers, using Logistic Regression, Decision Tree, Random Forest, SVM, and XGBoost classifiers- and utilizes stacked ensembles to bundle the outputs keeping accuracy levels high and improving robustness, especially for similar cropping classes

Recommendation Delivery Layer
This layer delivers recommendations via a minimalist web interface, downloadable PDF, provides the predicted crops, normalized bar charts for N, P, K, pH, and rainfall, and a short text description of existing soil-weather conditions, along with concise recommendations for planting, irrigation, fertilizer application, and harvesting.

B. Learning structure

Complementary Strengths
Random Forest to represent the nonlinearity and is an optimal solution in the problem of noisy inputs. The XGBoost models (fine grained feature) interactions. SVM provides good margins between narrow classes. Logistic Regression provides explainable coefficients and a fixed base level.

Stacked Ensemble
A meta-learner is a synthesis of model outputs learning the combination factor you can rely on this model. More or less the amount of such in such a context, thus it decreases variation and causes the boundaries of decisions to become less sharpened on those pairs of the crop which are most difficult to part, and at the same time obliterating the extreme accuracy of pairs.

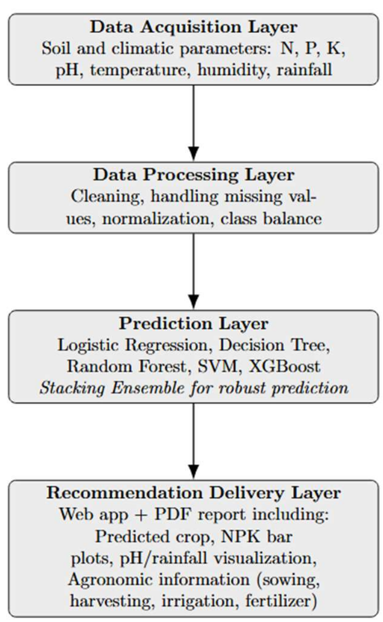

Figure 1. Proposed Crop Recommendation System Architecture

C. Scalability and deployment

Connectivity-aware Design
It is a webapp developed based on flask and would need a connection to the internet to start and use but will work via sluggish net or poor internet connections.

Modular Integration
As an input, it can be in the form of manual forms or plug-in sensors, and data are synchronized to the cloud to create data. Processing/sizing and new sensors/sites can be introduced without going back to the basic models.

Maintainability
We need to keep things by separating our work into different layers. This means we have one layer for getting information one for working with that information one, for making predictions and one for sharing the results. This approach makes it easy to update and check our system over time and make the long term maintenance system.

D. Figure reference

Figure 1 presents the complete workflow of the proposed system. The first one is the collection of soil and weather data (N, P, K, pH, temperature, humidity and rainfall). This is then processed and checked. Subsequently, crop prediction is carried out with the help of multiple machine learning models including a stacked ensemble. Lastly, the suggested crop and the associated data is provided via the web application with a generated PDF report.

IV. METHODOLOGY

The crop recommendation system was developed in five steps which included: data collection, data preparation, model training, performance measurement and the system deployment through a web application.

A. Data Collection

We have used the publicly available Crop Recommendation Dataset on Kaggle [19]. It consists of seven variables represented by 2200 records that include four soil properties, Nitrogen (N), Phosphorus (P), Potassium (K) and pH - and three weather variables, the rainfall (mm), temperature (°C), and humidity(%). These combined give a sufficient description of healthy soils and local weather to run a model of suggested crops.

B. Data Preparation

Under data validation, we tested the data to identify and remedy missing/abnormal values. Each of the soil and weather variables was put at close comparison. We also minded the class balance in such a way that the model would not favor more examples of crops that we had.

C. Model Training

We have trained several monitored models including the Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector machine (SVM) and Extreme Gradient Boosting (XGBoost) models. Not merely that, but then planked them together to bring any strong-points together. The data was divided into two 80/20-train dataset and test dataset, stratified sampling was used and his sample distribution of the classes in the datasets was preserved. To invalidate generalization, we used grid search and 5-fold stratified cross validation to randomise our hyper parameters. In each fold, four of which were utilized for training, and one which was utilized for validation, and we reported the mean accuracy and standard deviation across folds.

D. Performance Measurement

We assessed models on accuracy, precision, recall, and F1 - score. Confusion matrices helped to see which classes of crops confused most and where errors were concentrated. We also reported cross-validating means and standard deviations to include how stable models were across folds and we kept these values stable throughout to the Results section.

E. System Deployment

The system is a Flask-based web application that includes a lightweight web UI which is deployed on Render. The application produces a PDF report including information on the weather and soil, the predicted crops, normalized bar charts for N, P, K, pH and rainfall and brief operational instructions, including planting, harvest, irrigation, fertilizer schedules, etc. The pipeline is connectivity aware and it has been designed for live ingestion when attached to a set of sensors and/or external data services.

V. Results and Discussion

A. Model Effectiveness Measurement

The performance of the trained classifiers on the test data is summarized in table I. Overall, the predictive performances of all the models is high, indicating that soil and weather features (N, P, K, pH, rainfall, humidity and temperature) provide a good discriminative information for classification of crop. Logistic Regression and Decision Tree had a test accuracy of 97.95%, which led to that even rather simple learners are able to model a meaningful decision boundary for the given dataset. In the case of the single models, the best model was the Random Forest algorithm having the test accuracy of 99.55%, followed by a high precision, recall, F1-score. This good performance could be explained by the fact that it could capture the nonlinear relationships and feature interactions and also diminish overfitting via bootstrap aggregated and ensemble averaging. SVM gave 98.86% test accuracy which shows the SVM effectiveness of difference closely grown crop classes by maximizing the margins. XGBoost also showed a very good performance with 99.32% of test accuracy in modeling fine grained interaction between input features. The stacking ensemble achieved a test accuracy of 99.32% which indicates that the ensemble of complementary learners can be used to give stable performance while being competitive with the best performing individual model. This is in favour of ensemble learning in the situation of more important robustness than marginal increase maximum accuracy. Feature importance analysis using Gini index Random Forest shows that the nitrogen (N = 0.24) and phosphorous (P = 0.19) are most important factor affecting the prediction outcome followed by the rainfall (0.15) and temperature (0.14). Soil pH (0.10) and potassium (K = 0.09), in some way are also important to decisions pertaining to crop suitability. Confusion matrix inspection shows little misclassification between crop classes with majority of the predictions located correctly. Cross validation results are suggestive of stable generalizations as well. Random Forest gets 99.55±0.22% and stacking ensemble gets 99.50±0.15% which means that variance of ensemble is slightly less throughout a set of folds indicating that the ensemble has achieved a better consistency given different data splits.

TABLE I: PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS ON TEST SET

Model	Test Accuracy	Test Precision	Test Recall	Test F1-score
Logistic Regression	97.95%	98.09%	97.95%	97.95%
Decision Tree	97.95%	98.06%	97.95%	97.94%
Random Forest	99.55%	99.57%	99.55%	99.55%
SVM	98.86%	98.97%	98.86%	98.86%
XGBoost	99.32%	99.34%	99.32%	99.32%
Stacking Ensemble	99.32%	99.35%	99.32%	99.32%

B. Comparative Analysis

TABLE II: COMPARATIVE STUDY

Ref	Method & Dataset	Accuracy
[1] Optimizing Crop Yield (2025)	IoT-based crop recommendation using RF and ensemble learning, soil sensor data (2200 samples)	99.5%
[2] Optimizing Crop Selection (2025)	Deep learning–based crop recommendation using multi-feature soil data (CNN, ANN)	97–98%
[3] Unified Forecasting Models (2024)	Rainfall-based crop recommendation using SVR and AdaBoost	90% (R^2=0.90)
[4] Smart Crop Recommendation (2024)	Web-based crop recommendation using deep learning and soil testing	95%
[5] AgriFusion: Crop Recommendation (2024)	Ensemble-based crop recommendation using ML models (RF, SVM, LR, DT, XGBoost) on soil and weather data	94.48%
Proposed Work (RF)	Proposed work (Random Forest) using Kaggle soil and weather dataset	99.55% (test), 99.59% (CV)
Proposed Work (Ensemble)	Proposed work (Stacked Ensemble) using Kaggle soil and weather dataset	99.32% (test), 99.55% (CV)

A comparative study with respect to recent crop recommendation research is shown in Table II. While several previous works have achieved good results based on IoT-based systems, deep learning and hybrid models, the proposed model is able to show a competitive performance on the publicly available Kaggle Crop Recommendation dataset. While we found Random Forest had the highest single model test accuracy (99.55%), the stacking framework of ensemble classifiers was proposed to ensure robustness in the face of possible shifts of the deployment environment (e.g. sensor noise, regional differences of soil and weather). The ensemble makes use of complementary learning behaviour in between tree-based models, margin-based classifiers, and gradient-boosted learners to have better stability for challenging or overlapping crop classes. In deployment scenarios, Random Forest is always a great option for instances when maximum accuracy is desired while the ensemble approach is the preferred method of choice for scenarios when prediction consistency and robustness under varying conditions is more important.

C. System Outputs

The developed system offers crop recommendation and practical and easy to use outputs. The web application shows soil and weather parameters (N, P and K, pH, rain, humidity and temperature), the predicted crop, and some simple agronomic recommendations (when and time to sow, how and time to irrigate, fertiliser). Additionally, the system produces downloadable reports in PDF format in the form of normalized bar charts of the key parameters and a short textual summary of the field conditions to support the interpretability and transparency of the system.

D. Practical Implications

The proposed crop recommendation system functions as a data-driven decision supporting and reduces trial-and-error of the crop choice. By using machine learning-based predictions using the soil nutrient analysis and weather, the system can direct the farmer toward better crops that support for better yield potential and for better use of natural resources. The modular nature also provides the ability to integrate IoT sensors in the future to give automated real-time monitoring of the soil in a scalable manner even under low connectivity situations.

VI. CONCLUSION

The crop recommendation model proposed in this paper shows a good performance with integration into a deployable decision support system using multiple machine learning algorithms which include Random Forest, Logistic Regression, XGBoost, Decision Tree as part of ensemble approach. The proposed framework has demonstrated the ability to accurately predict the results of the crop whilst maintaining the recommendations for different types of crops. Random Forest in testing done best but ensemble method brought some advantages due to its capabilities of increasing the performance and reliability of system to be used for different applications. The system is created to function in an effective manner in the actual farming environment which exists in regions, that have limited access to internet services. The design structure of the system enables to change the agriculture requirements for organizations by adding new data sources and machine learning models and sensor inputs. The system provides scalable data-based solutions which allows the presence of the users to make the right choices for ventilation of the crops and resources management and to promote environmentally friendly farming methods. The research makes the connection between machine learning models and the requirements of agriculture that greatly promotes the development of precision agriculture technology.

FUTURE WORK

From future development point of view, the features will be conducted and added to the model to make it a complete smart agriculture system:

- IoT Integration: Using the real-time sensing from the soil and environmental sensors to give the farmers smart recommendations that depends on the existing condition.
- Climate Adaptation: Creating models that help farmers plan their crops better especially when the weather is unmanageable or extreme.
- User Accessibility: Providing the support of several languages and voice recognition to simplify it. platform on which various groups of farmers can access.
- Remote Sensing: Remote sensing refers to use of photographs captured on different devices such as the drones and satellites to monitor extensive regions, areas of land and harvests that would be better and more fruitfully cultivated.
- Scalability and Collaboration: Construction of networks in which the researchers and the farmers can share information, which will help in the improved predictions and the successful functioning of the system in other farming communities.

All these improvements will incorporate factors of machine learning, real-time information, available equipments and more efficient support structures to introduce a whole package of smart farming.

REFERENCES

[1] P. S. and V. K., "Optimizing crop yield—An IoT- and ML-based soil testing and crop recommendation system," in Proc. 3rd Int. Conf. Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 789–795, doi: 10.1109/IDCIOT64235.2025.10915073.

[2] V. S. Chowdam, P. S. Harsha Reddy, S. Likhitha, S. Nazma, and P. E. Sai, "IoT-enabled smart soil monitoring system with crop and fertilizer recommendations," in Proc. Int. Conf. Visual Analysis and Data Visualization (ICVADV), Tirunelveli, India, 2025, pp. 412–419, doi: 10.1109/ICVADV63329.2025.10961474.

[3] P. Kruthika, K. Srinidhi, M. Sowmithra, S. V. Karthick, and A. S. Prasad, "Unified forecasting models: Crop recommendation using predicted rainfall with machine learning," in Proc. 4th Int. Conf. Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 1111–1116, doi: 10.1109/ICUIS64676.2024.10866035.

[4] K. Kethineni, S. H. Mekala, M. Kodali, V. V. Kota, and J. P. Jampani, "A web-based agriculture recommendation system using deep learning for crops, fertilizers, and pesticides," in Proc. Int. Conf. Computational Intelligence for Green and Sustainable Technologies (ICCIGST), Vijayawada, India, 2024, pp. 1–6, doi: 10.1109/ICCIGST60741.2024.10717535.

[5] D. B. Sree et al., "AgriFusion: Crop recommendation system using ensemble learning methods," in Proc. 4th Int. Conf. Intelligent Technologies (CONIT), Karnataka, India, 2024, pp. 1–8, doi: 10.1109/CONIT61985.2024.10626639.

[6] M. Madhumitha and R. Ambikapathy, "Soil analysis and crop recommendation using deep learning," Int. Res. J. Modern Engineering Technology and Science, vol. 6, no. 8, pp. 1–8, Aug. 2024, doi: 10.56726/IRJMETS60753.

[7] S. B. Barshe et al., "Comprehensive analysis of artificial intelligence-based crop recommendation and soil analysis," in Proc. 2nd Int. Conf. Data Science and Information Systems (ICDSIS), Navi Mumbai, India, 2024, pp. 1–8, doi: 10.1109/ICDSIS61070.2024.10594125.

[8] M. R. Madhumathi et al., "Soil nutrient prediction and crop recommendation system," in Proc. Int. Conf. Circuit, Power and Computing Technologies (ICCPCT), Coimbatore, India, 2023, pp. 1–8, doi: 10.1109/ICCPCT58313.2023.10245685.

[9] A. J. Swetha et al., "Advanced soil fertility analysis and crop recommendation using machine learning," in Proc. 7th Int. Conf. Trends in Electronics and Informatics (ICOEI), Vijayawada, India, 2023, pp. 1035–1039, doi: 10.1109/ICOEI56765.2023.10125748.

[10] R. A. C. Ramachandra et al., "Crop recommendation using machine learning," in Proc. Int. Conf. Data Science, Networking and Security (ICDSNS), Bangalore, India, 2023, pp. 1–8, doi: 10.1109/ICDSNS58469.2023.10245154.

[11] P. A. Barvin and T. Sampradeepraj, "Crop recommendation systems based on soil and environmental factors using graph convolutional neural networks: A systematic literature review," in Proc. 10th Int. Electronic Conf. on Sensors and Applications (ECSA-10), 2023, pp. 1–10, doi: 10.3390/ecsa-10-16010.

[12] P. Kathiria, U. Patel, S. Madhwani, and C. S. Mansuri, "Smart crop recommendation system: A machine learning approach for precision agriculture," in Machine Intelligence Techniques for Data Analysis and Signal Processing, vol. 997, 2023, doi: 10.1007/978-981-99-0085-5_68.

[13] A. Sharma, M. Bhargava, and A. V. Khanna, "AI-Farm: A crop recommendation system," in Proc. 10th Int. Conf. Advanced Computing and Communication (ICACC), Chembur, India, 2021, pp. 1–8, doi: 10.1109/ICACC-202152719.2021.9708104.

[14] A. K. Anguraj et al., "Crop recommendation on analyzing soil using machine learning," Turkish Journal of Computer and Mathematics Education, vol. 12, no. 6, pp. 1784–1791, Apr. 2021.

[15] S. Ujjainia, P. Gautam, and S. Veenadhari, "A crop recommendation system to improve crop productivity using ensemble technique," Int. J. Innovative Technology and Exploring Engineering (IJITEE), vol. 10, no. 4, pp. 145–150, Feb. 2021.

[16] J. Raut and S. Mittal, "Soil fertility and crop recommendation using machine learning and deep learning techniques: A review," Turkish Journal of Computer and Mathematics Education, vol. 11, no. 2, pp. 1119–1127, 2020.

[17] B. K. R et al., "Crop recommendation system for precision agriculture," Int. J. Computer Science and Engineering, vol. 7, no. 5, pp. 1277–1282, May 2019.

[18] N. H. Kulkarni et al., "Improving crop productivity through a crop recommendation system using ensembling technique," in Proc. 3rd Int. Conf. Computer Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 1–6.

[19] A. Ingle, "Crop recommendation dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.