International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 11 Special Issue 02 | 2026

Machine Learning Based Smart Crop Disease Detection with Farm Advisory

Rahul Gajanan Patil[1], Vaishnavi Babarao Shinde[2], Ujjwala Narayan Tekade[3], Vijaya Ramdas Dhoke[4], Prof. Y. B. Jadhao[5]
[1,2,3,4]Student, Department of Computer Science and Engineering, Padmashri Dr.V.B. Kolte College of Engineering, Malkapur, Maharashtra, India
[5]Professor, Department of Computer Science and Engineering, Padmashri. Dr.V.B. Kolte College of Engineering, Malkapur, Maharashtra, India

DOI: 10.5281/zenodo.19251708

ABSTRACT

Agriculture is very important to the economy of several countries, including India, and it has challenges that include climate variability, soil deterioration problems, pest control issues, and market uncertainties. In this study, we introduced an intelligent smart farm platform which integrates crop recommendation, plant disease detection, weather prediction, fertilizer advice, market forecasts and agricultural counselling with machine learning and deep learning methods. The system proposed is web based which suggests real-time crop recommendation on the basis of soil nutrients, temperature, humidity, PH value and rainfall. Various machine learning methods (such as Decision Tree, Naive Bayes, Support Vector Machine, Logistic Regression, Random Forest) including KNN and XGBoost were developed, and the Random Forest method yield a highest accuracy for crop recommendation. Furthermore, a Plant Disease Identification model based on Convolutional Neural Network (CNN) has been developed to effectively identify and classify plant diseases from leaf images which can be used for early diagnostics and intervention. The supplementary modules such as weather forecasting, fertilizer recommending and market price prediction add on to make the farmers work smarter by supporting them in their decision-making process. It enhances agricultural productivity, and minimizes crop losses through pest management, sustainable and technology driven farming practices.

Keywords:- machine learning, Smart agriculture, crop recommendation, plant disease identification, machine learning, convolutional neural network, random forest.

I. INTRODUCTION

The use of machine learning and artificial intelligence (AI) has brought significant changes across many fields, with agriculture emerging as one of the major beneficiaries. Among the various applications of these technologies, smart crop recommendation and plant disease identification have become especially important. In recent years, machine learning and data science have transformed traditional farming practices by enabling data-driven decision- making. One notable advancement is the development of intelligent crop recommendation systems that use predictive models to guide farmers in selecting crops best suited to their soil and environmental conditions, along with early detection of crop diseases using image-based techniques. In this paper, we present a comprehensive Smart Crop Recommendation System developed using Python and multiple machine learning algorithms for classification tasks. The proposed system supports farmers by analyzing soil properties, climatic conditions, and geographical factors to help them make informed farming decisions. Users can input essential parameters such as soil nutrient levels (nitrogen, phosphorus, and potassium), rainfall, temperature, humidity, pH value, and geographical location. Based on a large dataset and the application of several classification models, the system predicts the most suitable crops for cultivation in a specific region.

In addition to crop recommendation, the system includes a Plant Disease Identification module that uses image classification techniques. Farmers can upload images of plant leaves showing disease symptoms, and a Convolutional. Neural Network (CNN) is employed to accurately identify and classify plant diseases. This enables early detection and timely intervention, helping to reduce crop losses and improve yield quality. The system also offers several supporting features, including weather forecasting, fertilizer recommendation, market price prediction, smart farming guidance, agricultural learning resources, information on government schemes and insurance, and an Agri support hub. These modules work together to provide comprehensive assistance to farmers. The entire system is deployed through a user-friendly web interface developed using Streamlit, ensuring easy access and smooth interaction. Through this platform, users can enter data effortlessly, receive intelligent recommendations, and gain actionable insights related to crop selection, crop health, and overall farm management.

Artificial intelligence is increasingly being adopted in fields such as educational technology, management sciences, and operational research. Intelligence is commonly defined as the ability to acquire and apply knowledge to solve complex problems. As AI systems continue to evolve, intelligent machines are expected to perform tasks that traditionally required human effort. Artificial intelligence focuses on the development of computer systems and software capable of reasoning, learning, acquiring knowledge, communicating, manipulating data, and recognizing objects, making it a powerful tool across multiple domains.

II. LITERATURE REVIEW

The rapid growth of smart farming technologies has significantly changed modern agricultural practices. These technologies offer practical solutions to improve crop productivity, make better use of resources, and manage plant diseases more effectively. Recent studies emphasize the need for intelligent decision-support systems that combine crop recommendation, disease detection, and additional agricultural services. Early detection of crop diseases allows farmers to take timely preventive measures, which helps control disease spread and reduces overall crop losses [11].

A. Machine Learning Algorithms

The application of machine learning has brought major improvements to agriculture by enabling accurate crop prediction and better farm management. This section discusses key machine learning algorithms that are commonly used to predict crop yield and recommend suitable crops for cultivation. By analyzing agricultural data, these algorithms contribute to improved productivity and long-term sustainability [10].

Decision Trees

Decision trees are widely used machine learning algorithms known for their effectiveness in both classification and regression tasks [3]. In agricultural applications, decision trees help model the relationship between soil characteristics, environmental conditions, and expected crop yield. By dividing data based on input features such as soil pH, moisture content, and temperature, decision trees provide a clear and interpretable decision-making structure. This transparency makes them especially useful for farmers and agricultural experts who need understandable and actionable insights.

Random Forests

Random forests extend the concept of decision trees by using an ensemble learning approach that combines multiple decision trees to improve prediction accuracy. This technique reduces overfitting by averaging the outputs of many trees, resulting in more stable and reliable predictions [6]. Random forests are particularly effective when working with large agricultural datasets that include multiple factors influencing crop growth and health. Their ability to handle complex data makes them well suited for accurate crop yield prediction.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are a specialized class of neural networks designed for image processing tasks. Inspired by the human visual system, CNNs are widely applied in agriculture for plant disease identification. By analyzing images of crop leaves, CNNs can automatically learn disease-related patterns and accurately classify different plant diseases. Their capability to extract meaningful features from images makes CNNs highly effective for analyzing field images and data collected through remote sensing technologies.

B. Optimization of Web-Based Crop Recommendation Systems

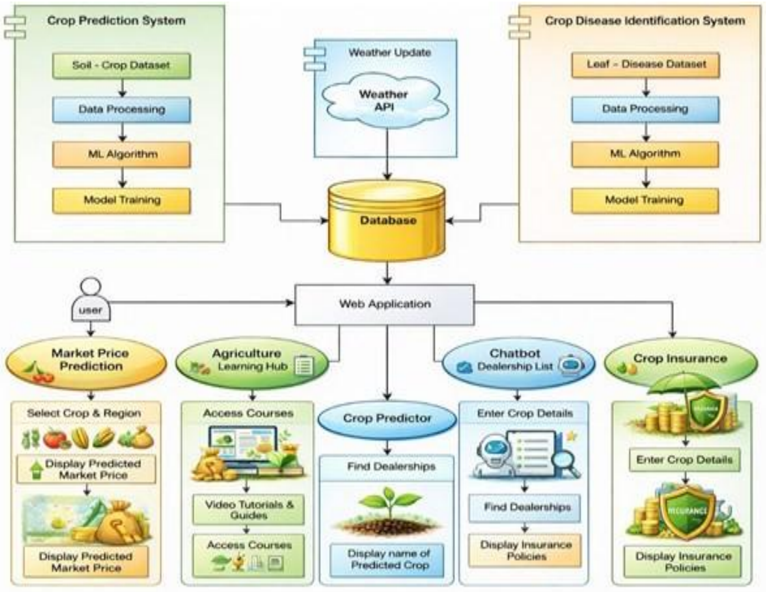

Fig. 1. System architecture

For web-based smart farming applications, optimizing both system performance and user experience is essential to ensure widespread adoption among farmers and agricultural stakeholders. Modern platforms aim to integrate multiple functional modules, including crop recommendation, plant disease identification, weather forecasting, fertilizer guidance, market price prediction, and agricultural support services, into a single system. A responsive and user- friendly interface allows farmers to access the system easily across various devices such as smartphones and desktop computers, as shown in Fig. 1. Fast response times and minimal processing delays are especially important for users in rural areas with limited internet connectivity. Incorporating real-time external data sources, including weather updates and market information, further improves the accuracy of recommendations. In addition, the inclusion of educational resources, government schemes, insurance details, and expert support services enhances overall usability and farmer engagement, making web-based smart farming systems more effective and practical.

III. PROPOSED STUDY

The implementation phase plays a crucial role in the overall framework of the Smart Crop Recommendation and Plant Disease Identification system. This phase focuses on converting the theoretical concepts and methodologies discussed in earlier sections into a practical and functional system [10]. It involves systematic system development, thorough execution of testing procedures, and detailed analysis of results to ensure that all specified requirements are met. The following subsections describe the adopted methodologies, testing and validation approaches, result analysis, and quality assurance measures implemented to ensure the reliability and effectiveness of the system [8].

A. Methodology

Datasets

The crop recommendation dataset consists of a total of 2,200 records, with each record containing eight attributes: Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH, Rainfall, and Label. The NPK values represent the nutrient content of the soil, while temperature, humidity, and rainfall indicate the average environmental conditions. The pH value reflects soil acidity or alkalinity. The label column represents the crop type that is most suitable for cultivation under the given conditions and serves as the target variable for prediction. For plant disease identification, a secondary dataset is used, comprising 70,295 plant images that display a wide range of diseases. The dataset occupies approximately five gigabytes of storage and consists of images standardized to a resolution of 128 × 128 pixels. It includes 38 different classes, covering 14 plant species and 26 distinct diseases. After training multiple models, the algorithm achieving the highest accuracy is selected for further evaluation and deployment.

Data Preprocessing

Data preprocessing is a critical step that involves transforming raw data into a suitable format for analysis and model training. In the plant disease identification module, the raw images obtained from the dataset may contain noise and inconsistencies. Therefore, preprocessing techniques such as resizing, normalization, and noise reduction are applied before feeding the images into the learning model. These steps help improve model performance and prediction accuracy.

Train and Test Split

The dataset is divided into training and testing subsets using the train-test_split () function available in the scikit-learn library. Out of the 2,200 data records, 80% (1,760 samples) are allocated for training the model, while the remaining 20% (440 samples) are reserved for testing. This split ensures that the model is trained on sufficient data while allowing reliable evaluation of its performance on unseen data.

Model Prediction

In machine learning, prediction refers to the output generated by a trained model when it is applied to new or unseen data. The model predictions are obtained using the predict () method, which processes the test feature dataset and returns an array of predicted values [9].

For plant disease identification, model evaluation follows a structured approach:

- Initially, 80% of the image dataset is used for training, while the remaining 20% is reserved for validation.
- Validation data is used to evaluate model accuracy by applying the prediction function and extracting relevant features.
- Additional plant images are captured and tested to verify disease detection accuracy once optimal results are achieved during validation.

2. Label Confusion Matrix and Classification Report

The strategies of Confusion Matrix and Classification Report are imported from the metrics module within the scikit learn library. They are computed via using the actual labels present in the check datasets along the expected values [1]. The Confusion Matrix affords a comprehensive breakdown of the frequency of real negatives, false negatives, real positives, and false positives within the dataset [1].

Precision denotes the capability of a classifier to correctly pick out the wide variety of fine predictions among the ones which might be deemed fantastic which is shown in "(1)". It is computed because the ratio of real positives to the sum of genuine positives and fake positives for every class [1]. Where the accuracy of Precision-Positive Predictions; True Positive (TP) and False Positive (FP)

Precision = TP/ TP+FP (1)

Recall, in the context of classification, signifies the classifier's ability to correctly identify all high-quality times from the confusion matrix which is shown in equation. Mathematically, it's miles computed because the quotient of real positives divided through the sum of proper positives and fake negatives for every magnificence which is shown in "(2)".

Recall = TP/TP+FN (2)

Where recall is the proportion of correctly detected positives; False Negative, or FN The F1-Score rating represents a weighted harmonic suggest of precision and don't forget, in which a price of zero. 0 denotes the poorest overall performance, at the same time as 1.0 signifies most efficient performance. As precision and do not forget are imperative additives in its calculation, F1 ratings commonly yield lower values as compared to accuracy measurements which is shown in "(3)".

F1−Score = 2∗PR/(P+R) (3)

Where P-Precision; R-Recall

3. Accuracy

Model accuracy indicates the proportion of correct predictions relative to the total number of predictions made. It is calculated using the accuracy score () function from the scikit-learn metrics module and is expressed as shown in Equation (4):

Accuracy = TP+TN/TP+TN+FP+FN Where TP-True Positive; FP-False Positive; TN-True Negative; FN-False Negative

B. Integration with Smart Farming Features

All trained models and processed datasets are stored in a centralized database and accessed through a web application developed using Streamlit. The application offers a range of features, including crop recommendation, plant disease identification, real-time weather updates, fertilizer recommendations, market price prediction, smart farming guidance, agricultural learning resources, government schemes and insurance information, and an Agri support hub. This integrated system provides farmers with comprehensive support through a single, user-friendly platform, enabling informed decision-making and improved farm management.

IV. RESULT ANALYSIS

We have chosen seven classification algorithms and checked how accurate each one is. For the second task, the best convolutional neural network (CNN) architectures for classifying photographs were used. The ensuing tables gift the corresponding outcomes.

able.1 offers the accuracy of crop advice responsibilities across diverse algorithms [4]. This desk unequivocally demonstrates that the random wooded area set of rules surpasses all others, accomplishing an extraordinary accuracy of 99.54%.

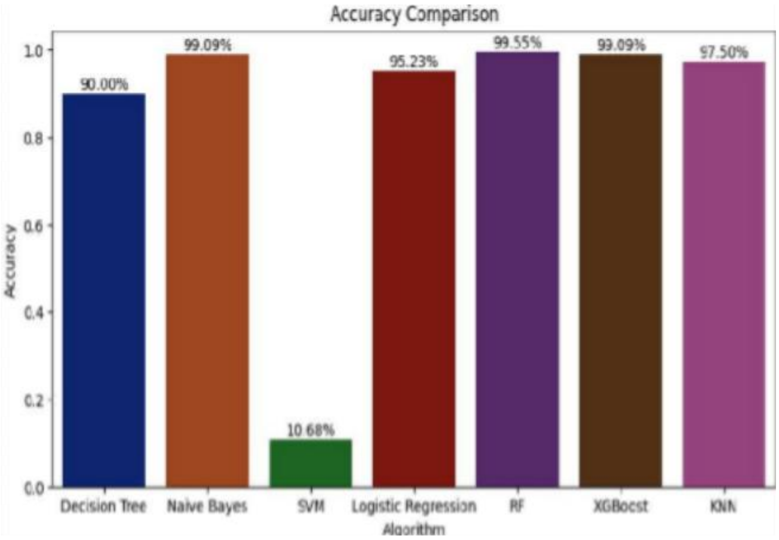

Fig. 2. Accuracy comparison

TABLE I. ACCURACY VS ALGORITHMS

Algorithm	Accuracy
Decision Tree	90.0
Gaussian Naive Bayes	99.09
Support Vector Machine (SVM)	10.68
Logistic Regression	95.23
Random Forest	99.55
XGBoost	99.09
KNN	97.50

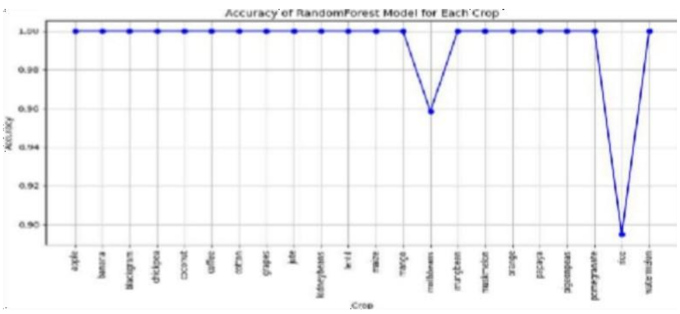

Figure. 3. Accuracy vs Crop graph for random forest model

Figure 3 illustrates that the Random Forest model delivers consistently strong performance across different crop types often outperforming other algorithms. While certain models achieve higher accuracy for specific crops, their performance tends to vary across conditions. In comparison, Random Forest shows greater stability and adaptability, making it a reliable option for crop prediction under diverse agricultural scenarios. In this study, a

Convolutional Neural Network (CNN) was employed for plant disease detection using a dataset of 17,572 images representing 38 classes of plant diseases and healthy plant conditions. The dataset was prepared using the tf. keras.utils.image_dataset_from_directory function, with all images resized to 128 × 128 pixels. class labels were automatically derived from the directory structure. A pre-trained CNN model, loaded from a. keras file, was used for disease prediction [5], and dropout layers were incorporated to minimize overfitting. For testing purposes, an image of a potato plant affected by early blight was processed using OpenCV. The image was converted from BGR to RGB format, resized, and then passed to the trained model. The final prediction was obtained using np. argmax on the output probability distribution. These results highlight the effectiveness of CNNs in automated plant disease detection, enabling early diagnosis and improved crop management. Figure 4 shows a sample image of a diseased leaf used in this study. The presence of visible features such as spots, discoloration, and other disease symptoms allows the model to learn meaningful patterns and accurately identify different plant diseases. This contributes to smarter crop management practices and supports healthier crop.

Figure. 4. Example of diseased image

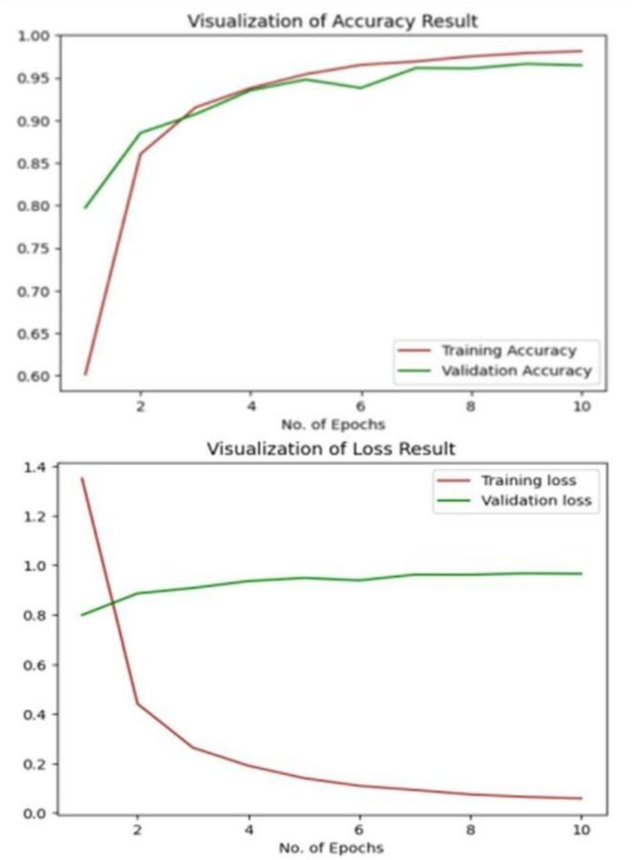

Figure. 5. Training and Validation (accuracy and loss) simple CNN

The graph illustrates the performance of the CNN-based plant disease detection model through training and validation accuracy and loss. High accuracy indicates effective classification of healthy and diseased plants, while low loss reflects good learning and generalization to unseen data [5][12]. Learning curves help analyze the model's training behavior over time. As shown in **Figure. 5**, both training and validation losses decrease steadily with minimal fluctuation, and the gap between them remains small, indicating good model stability.

Similarly, training and validation accuracy improve consistently. These results demonstrate strong fitting and reliable performance of the proposed CNN model [2].

V. CONCLUSIONS AND FUTURE SCOPE

In conclusion, this project successfully developed a smart crop recommendation and disease detection system using machine learning and real-time data processing. The system enhances crop management by providing accurate insights and early disease identification. Its modular, scalable design and user-centric interface ensure reliability, ease of use, and accessibility for farmers. Overall, the integration of advanced technologies supports informed decision-making and promotes sustainable agricultural practices.

VI.REFERENCES

[1] M. Shripathi Rao, A. Singh, N. V. Subba Reddy, and D. U. Acharya, "Crop prediction using machine learning," Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012033, Jan. 2022.

[2] Md. Manowarul Islam et al., "DeepCrop: Deep learning-based crop disease prediction with web application," Journal of Agriculture and Food Research, vol. 14, pp. 100764–100764, Dec. 2023.

[3] Kar, N. Nath, U. Kemprai, and Aman, "Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection," International Journal of Communications, Network and System Sciences, vol. 17, no. 2, pp. 11–29, Feb. 2024.

[4] G. Rani *et al*., "Automated crop and fertilizer disease prediction using CNN," in *Proc. 2nd Int. Conf. Adv. Computing and Innovative Technologies in Engineering (ICACITE)*, Apr. 2022.

[5] Tirkey, K. K. Singh, and S. Tripathi, "Performance analysis of AI- based solutions for crop disease identification, detection, and classification," Smart Agricultural Technology, vol. 5, p. 100238, Oct. 2023.

[6] K. Ramu and K. Priyadarsini, "A Review on Crop Yield prediction Using Machine Learning Methods," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Oct. 2021.

[7] Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images," Cluster Computing, Aug. 2022.

[8] K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, "Crop Yield Prediction using Machine Learning and Deep Learning Techniques," Procedia Computer Science, vol. 218, pp. 406–417, 2023.

[9] *M. Shripathi Rao, A. Singh, N. V. Subba Reddy, and D. U. Acharya, "Crop prediction using machine learning,"* Journal of Physics: Conference Series*, vol. 2161, no. 1, p. 012033, Jan. 2022.*

[10] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla, and K. Shaurya, "Crop Recommender System Using Machine Learning Approach," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Apr. 2021.

[11] Raju, A. D. V., and A. P. B. V., "CropCast: Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction," *Kuwait Journal of Science*, p. 100160, Dec. 2023.

[12] P. Chauhan, H. L. Mandoria, A. Negi, and R. S. Rajput, "Plant Diseases Concept in Smart Agriculture Using Deep Learning," in *Advances in Environmental Engineering & Green Technologies* book series, pp. 139–153, Oct. 2020.