



Design And Implementation Of An Integrated System For Crop Recommendation And Leaf Disease Detection Using Ensemble Learning

Pilla Divya Vidhyasree, Lalam Bhavana, Subhadip Mohanta, Gorja Govardhan

B. Tech Students, Department of CSE (Artificial Intelligence and Machine Learning), Dadi Institute of Engineering and Technology, NH-16, Anakapalle, Visakhapatnam-531002, A.P

Mr. A. Venkateswara Rao -Department of CSE (Artificial Intelligence-Machine Learning) Diet, Anakapalle, India.

Assistant Professor: Mrs.Chandrika, Department of CSE (Artificial Intelligence-Machine Learning) Diet, Anakapalle, India.

ABSTRACT

Agriculture underpins global food security and economic stability but faces challenges such as plant diseases and inefficient crop selection, which threaten productivity and sustainability. This study proposes an integrated system combining crop recommendations and leaf disease detection using advanced deep-learning techniques and ensemble learning. The plant disease detection module employs an ensemble of DenseNet201, VGG19, and InceptionV3 to accurately classify and diagnose plant diseases from leaf images, thereby leveraging the strengths of these architectures for feature extraction. Simultaneously, the crop recommendation module analyzes environmental factors such as soil properties, climate conditions, and geographical data to suggest optimal crops using machine learning models such as Gaussian NB. Comprehensive evaluations demonstrate that the proposed system achieves a high classification accuracy for both tasks, making it robust and reliable for real-world agricultural scenarios. This unified platform empowers farmers by providing actionable insights to enhance productivity, reduce crop losses, and promote sustainable agricultural practices. This system leverages AI-driven solutions to integrate traditional farming practices with modern technological advancements and empower sustainable agriculture.

Keywords: Smart Agriculture, Crop Recommendation, Plant Disease Detection, Ensemble Learning, Deep Learning, DenseNet201, VGG19, InceptionV3.

INTRODUCTION

Agriculture also plays a pivotal role in ensuring food security and economic stability. However, farmers often face critical challenges that affect agricultural productivity, including plant diseases and suboptimal crop selections. Plant diseases can lead to significant yield losses and reduced crop quality, whereas poor crop selection based on generalized practices often results in inefficient resource utilization and lower profitability. Addressing these issues is essential for sustainable farming and the long-term viability of agriculture. Traditional methods of disease detection and crop recommendation are labor-intensive and requires expert knowledge, which may not always be accessible, especially in rural areas. Manual inspection of plant diseases is prone to errors, time consumption, and heavy reliance on human expertise. Similarly, crop selection decisions often overlook crucial factors such as soil composition, climatic conditions, and historical crop rotation patterns, leading to suboptimal choices. Artificial intelligence (AI) and machine learning (ML) offer promising solutions to these challenges by enabling automated, accurate, and efficient agricultural decision-making systems. Agriculture is a leading profession in developing nations such as India. Crop losses significantly impact national economies, particularly agriculture-dependent regions. Plant diseases are a significant cause of crop losses. They may degrade crop quality and reduce productivity. Famines and increased unemployment in the agricultural sector are possible outcomes. A lack of farmer awareness poses a significant challenge to the timely identification and prevention of plant diseases, often leading to substantial crop losses. Various internal and external factors generally infect plants. Internal factors contributing to plant infections include the presence of pathogens such as fungi, viruses, and molds, and external factors encompassing environmental conditions, such as precipitation, temperature, and humidity, which may sporadically trigger their occurrence. These diseases can have serious financial consequences and affect the livelihoods of farmers. Plant diseases seriously harm crops and cause food shortages, increasing human food sustainability and security. The accurate and timely identification of plant diseases is essential for effective crop management and disease control. Modern image processing and the recognition of pattern tools have found several ways to help farmers and agricultural specialists identify diseases. Various methods have been used to identify plant diseases. Machine learning (ML) and deep learning (DL) techniques can be employed to train models that leverage image data to detect plant diseases and crop recommendations. In recent years, several XXX-X-XXXX-XXXX-X/XX/\$XX.00 ©20XX IEEE DL principles have been applied in agriculture to address various problems, such as insect recognition, fruit identification, fruit disease identification, categorization of plant leaves, and leaf disease identification. The application of typical ML techniques for real-time disease identification is highly challenging. As a result, DL techniques can point developers to the right path to overcome these challenges and advance the agricultural industry. Different images of various plant parts can be collected to create a model that can detect plant diseases. Traditional ML techniques also need to be improved in real-world detection situations owing to their complicated, non-uniform backgrounds. Various illnesses can affect different crops and each disease may manifest differently. It is challenging to create a comprehensive system that can identify various illnesses in various crop species. Large agricultural fields should be manually checked for disease symptoms. To overcome these issues, a new DL method has been proposed for detecting and categorizing plant leaf diseases. For crop recommendation, a Gaussian Naive Bayes (Gaussian NB) classifier was used to analyze critical environmental factors, such as soil pH, nutrient levels, temperature, humidity, and rainfall, to recommend optimal crops for a given region. The Gaussian NB model, known for its simplicity and efficiency, is particularly suited for handling continuous features and providing probabilistic predictions, thus enabling farmers to make data-driven decisions. The proposed system was designed to operate as a unified platform that is accessible to farmers via a user-friendly interface. The platform allows farmers to upload leaf images for disease diagnosis and input environmental parameters for crop recommendation. By integrating these functionalities, the system

empowers farmers to make informed decisions, reduce crop loss, and optimize resource utilization. The main contributions of this study are as follows: An ensemble learning approach for plant leaf disease detection combines DenseNet201, VGG19, and InceptionV3 models to improve accuracy and generalization. A Gaussian NB-based crop recommendation module was tailored to environmental conditions for optimal crop selection. A unified, user-friendly system bridges the gap between traditional farming practices and modern AI-driven solutions, thereby promoting sustainable agriculture. Through extensive experimentation and evaluation, the system demonstrated its efficacy in addressing real-world agricultural challenges, making it a valuable tool for farmers aiming to enhance productivity and sustainability.

Literature Review: Agriculture has benefited significantly from the integration of artificial intelligence (AI) and machine learning (ML), particularly in crop recommendation and plant disease detection. This section explores the existing methodologies and their advancements.

Crop Recommendation Systems Traditional crop recommendation systems rely on generalized data or expert opinions, often overlooking critical environmental factors such as soil pH, temperature, and rainfall. Early models such as Support Vector Machines (SVM) and Decision Trees offered moderate success but struggled with scalability and accuracy in real-world applications. Probabilistic models like Gaussian Naive Bayes (Gaussian NB) have shown promise due to their ability to handle continuous features and provide robust crop predictions.

Plant Disease Detection Plant diseases can cause substantial agricultural losses, making accurate and timely detection of these diseases essential. Initial approaches used image processing techniques, including feature extraction and classification using an SVM and k-nearest neighbors (kNN). However, these methods required handcrafted features and lacked robustness against variations in disease patterns and image quality. The emergence of deep learning has revolutionized disease detection, with Convolutional Neural Networks (CNNs) such as VGG16 and ResNet50 achieving high accuracy on datasets such as PlantVillage. These models automatically learn features from images, thereby eliminating the need for manual feature engineering. More advanced architectures, such as DenseNet and InceptionV3, further improve performance by capturing hierarchical and multiscale features.

Ensemble Learning in Disease Detection Ensemble learning has gained traction in plant disease detection by combining multiple models to enhance accuracy and generalization. Recent studies demonstrated that ensembles of ResNet, VGG, and Inception architectures outperform individual models. For instance, integrating DenseNet and InceptionV3 has been shown to leverage complementary strengths, improving classification robustness and reducing overfitting.

Unified Systems Few studies have integrated crop recommendations and disease detection in a single platform. Existing systems typically address these tasks separately, thereby limiting their usability in real-world scenarios. Unified systems that combine deep learning for disease detection and machine learning for crop recommendation can provide farmers with holistic solutions. However, such systems are rare and often lack scalability and adaptability to diverse agricultural conditions.

Research Gap Current research often focuses on isolated tasks such as crop recommendation or disease detection, without combining them into a unified system. Moreover, challenges such as overfitting, limited datasets, and lack of user-friendly deployment remain unaddressed. This study aims to fill this gap by developing a robust, unified system that integrates Gaussian NB for crop recommendation and an ensemble of DenseNet201, VGG14, and InceptionV3 for plant disease detection.

METHODOLOGY

Proposed Methodology:

The proposed system is an integrated approach that combines crop recommendation and plant leaf disease detection. The system is designed to aid farmers and agricultural stakeholders by providing accurate recommendations for crops based on environmental factors and diagnosing plant diseases from leaf images using advanced machine learning techniques.

1. CROP RECOMMENDATION SYSTEM

Objective: The crop recommendation system aims to suggest the most suitable crops based on environmental and soil conditions, optimizing resource usage and ensuring high yield for the farmer.

Steps Involved:

a) Data Collection:

Features Used: The system gathers various environmental and soil-related data, including:

Temperature, Humidity, Soil pH, Rainfall, Soil texture and moisture content, Geographic location

b) Data Preprocessing:

The collected data is cleaned and processed to handle missing values, outliers, and noise.

c) Data normalization and scaling techniques are employed to standardize feature ranges, ensuring effective model convergence during training.

d) Model Selection and Training:

Model Used: Gaussian Naive Bayes (GaussianNB) is chosen for its simplicity, efficiency, and good performance with smaller datasets.

The dataset is split into training and testing sets.

e) Model Evaluation:

After training, the model is evaluated using the testing set.

Accuracy: The model achieved 99.8% accuracy, indicating that it performs excellently in predicting the most suitable crops for different conditions.

f) User Interface:

A user-friendly interface is created where farmers can input their environmental data, and the system will output crop recommendations.

Soil texture and moisture content Geographic location

g) Data Preprocessing:

The collected data is cleaned and processed to handle missing values, outliers, and noise.

h) Data normalization and scaling techniques are employed to standardize feature ranges, ensuring effective model convergence during training.

i) Model Selection and Training:

Model Used: Gaussian Naive Bayes (GaussianNB) is chosen for its simplicity, efficiency, and good performance with smaller datasets.

The dataset is split into training and testing sets. o The model is trained on the training set, where it learns the relationship between the environmental conditions and suitable crops.

j) Model Evaluation:

After training, the model is evaluated using the testing set. o Accuracy: The model achieved 99.8% accuracy, indicating that it performs excellently in predicting the most suitable crops for different conditions.

2. PLANT LEAF DISEASE DETECTION SYSTEM

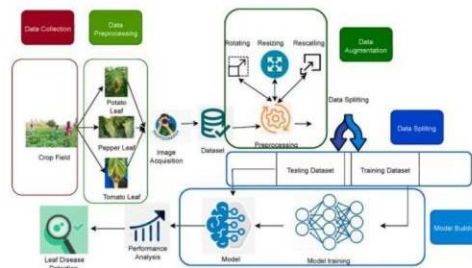


Fig. 1. Proposed work-flow diagram.

Objective: The plant leaf disease detection system aims to diagnose plant diseases from leaf images and provide appropriate disease labels, enabling farmers to take early actions and mitigate crop damage. The development of a comprehensive system for the detection of various diseases across diverse crop species is complex. Manual monitoring of large agricultural fields for disease symptoms is time consuming. To overcome these issues, a new DL method has been proposed for detecting and categorizing plant leaf diseases. The objective of this study was to develop and deploy deep learning models capable of accurately predicting rice leaf disease.

a) Data Collection:

- A large dataset of plant leaf images is used for training the model. The dataset includes images from various crop types (e.g., apple, tomato, etc.) and diseases (e.g., healthy, scab, black rot, etc.). High - quality plant leaf images are collected, ensuring representation of both healthy and diseased samples. The dataset is systematically partitioned into training, validation, and testing subsets to ensure robust model evaluation and parameter tuning. □
- The data is divided into three categories: Training Data: Used to train the models. Validation Data: Used to tune model parameters and evaluate intermediate performance. Testing Data: Used to evaluate the final model performance.
- Various plants and diseases have been collected from Plant Village dataset where total of approx. 70000 images dataset had been taken to train the model.

TABLE 2. The dataset's image count for each illness class.

Plant Name	Image Number
Apple	7771
Blueberry	1816
Cherry	3509
Corn	7316
Grape	7222
Orange	2010
Peach	3566
Pepper	3901
Potato	7641
Raspberry	1781
Soybean	2022
Squash	1736
Strawberry	3598
Tomato	18345

- Images are resized to a uniform dimension (e.g., 224x224 pixels) to match the input size requirements of each model.
- Pixel values are normalized to [0, 1] to standardize the input.



Fig. 3. Examples of Diseased image.

a) Data Preprocessing:

The images undergo preprocessing to enhance features and remove noise. Resizing: Images are resized to a fixed size to ensure uniformity.

Normalization: Pixel values are normalized to a range of 0 to 1.

Augmentation (optional): Techniques such as rotation, flipping, and cropping can be applied to increase the diversity of the dataset, enhancing the model's robustness.

b) Model Selection:

DenseNet201, VGG19, and InceptionV3 in Disease Detection This research employs an ensemble of DenseNet201, VGG19, and InceptionV3 models to leverage their unique strengths: 1. DenseNet201: DenseNet201, part of the DenseNet family, connects each layer to every subsequent layer in a feedforward fashion, allowing the reuse of features. This reduces the number of parameters and promotes efficient gradient flow, resulting in improved training and performance on smaller datasets. DenseNet's dense connections are particularly useful for plant disease detection, where the dataset may have variations in image quality and disease patterns.

2. VGG19: A simplified variant of the VGG family, VGG19 maintains the uniform convolutional structure of VGG16 but uses fewer layers, balancing computational efficiency and performance. It excels in capturing fine-grained details in plant leaf images, making it a robust choice for disease classification tasks.

3. InceptionV3: InceptionV3 introduces inception modules, allowing the model to capture multi-scale features by performing parallel convolutions with filters of different sizes (e.g., 1x1, 3x3, and 5x5). This multiscale feature extraction enhances the model's ability to identify subtle disease patterns across diverse datasets, improving overall classification robustness.

c) Model Training:

Each model is trained on the training dataset. The models learn to extract features from leaf images and classify them into disease categories.

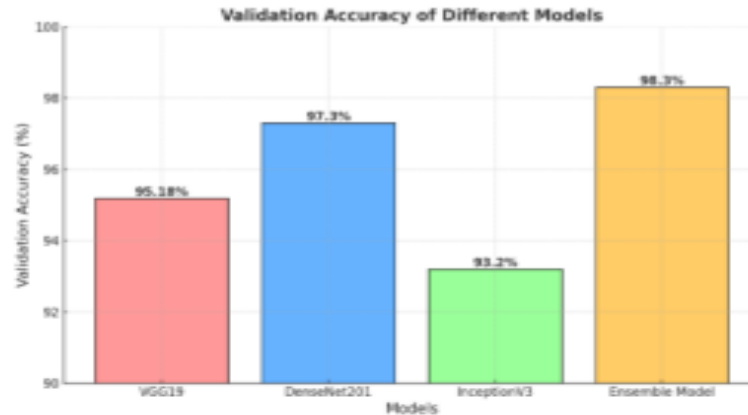
For each model: The training loss and accuracy are monitored during the process. The models are validated using the validation dataset to avoid overfitting and tune hyperparameters.

d) Ensemble Model:

After training the individual models (VGG19, DenseNet201, and InceptionV3), their outputs are combined in an ensemble model to leverage the strengths of each.

The ensemble model combines the predictions of the three models, providing more robust and accurate results than any individual model. . Performance of models

Model Name Training loss Training Accuracy Validation loss Validation Accuracy VGG19 0.0158 99.59% 0.1606 95.18% DenseNet201 0.0779 99.84% 0.099 97.3% Inception V3 0.1 98.8% 0.2469 93.2% Ensemble model 0.1 99% 0.18% 98.3%



Performance of models

Model Name	Training loss	Training Accuracy	Validation loss	Validation Accuracy
VGG19	0.0158	99.59%	0.1606	95.18%
DenseNet201	0.0779	99.84%	0.099	97.3%
Inception V3	0.1	98.8%	0.2469	93.2%
Ensemble model	0.1	99%	0.18%	98.3%

a) Model Evaluation:

Each model, including the ensemble, is evaluated using the testing dataset.

The evaluation is performed based on metrics like accuracy, precision, recall, and f1 -score.

The ensemble model achieved the highest accuracy of 98.3%, indicating that it outperforms individual models in real- world scenarios.

b) User Interface:

A web-based or mobile interface allows farmers to upload images of plant leaves.

The system processes the image, classifies the disease, and provides suggestions for treatment or preventive measures.

3. SYSTEM INTEGRATION

Objective: To create a unified system where crop recommendations and disease detection are combined into one user- friendly platform.

Steps Involved:

a) System Architecture:

The system is built with two major modules:

1. Crop Recommendation Module: Provides recommendations based on environmental data using the trained Gaussian NB model.
2. Leaf Disease Detection Module: Analyzes leaf images and detects diseases using the ensemble model of VGG19, DenseNet201, and InceptionV3.

b) Integration Process:

The two modules are integrated into one cohesive application.

Users can input data in one of two ways: Environmental Data: For crop recommendations. Leaf Images: For disease detection.

c) Output:

Crop Recommendation: Based on the input environmental data, the system provides accurate crop suggestions.

Disease Detection: Based on the uploaded leaf image, the system identifies the disease and provides appropriate recommendations.

d) Testing and Validation:

The integrated system is tested with real-world data to ensure the reliability of the crop recommendation and disease detection processes.

The user-friendly solution is convenient to use and able to detect leaf diseases. Feedback from initial users is incorporated to improve the interface and functionality.

4. RESULTS AND PERFORMANCE EVALUATION

Crop Recommendation Accuracy: The Gaussian Naive Bayes (Gaussian NB) model demonstrates exceptional performance, achieving 99.8% accuracy in crop recommendation across diverse environmental conditions.

The ensemble model, which combines the outputs of VGG19, DenseNet201, and InceptionV3, provides the highest performance with 98.3% validation accuracy, making it highly effective for detecting plant diseases. The model shows exceptional results by detecting many plants leaf disease detection and achieved 100% accuracy for many plants like Apple_Black_rot, Apple___Cedar_apple_rust, Apple_healthy,etc... Ensemble model combines the outputs of 3 accurate models to get the better combined solution. The ensemble model, which combines the outputs of VGG19, DenseNet201, and InceptionV3, provides the highest performance with 98.3% validation accuracy, making it highly effective for detecting plant diseases The ensemble approach improves the performance by using a votin g mechanism or averaging the outputs from the individual models

Evaluation Metrics: We employ performance measures including Accuracy, Precision, Recall and F1 Score to measure the performance of our method. The confusion matrix is utilized to determine the accuracy, precision, recall and f1 - score.

Confusion Matrix: It is an $N \times N$ matrix which is used for evaluating the performance of a classification model where N is the number of classes. It compares the actual labels to the model's predictions and gives us a holistic view regarding the performance of our model along with the type of errors it is making. The $N \times N$ matrix consists of two kinds of values: positive and negative. The columns and rows of the matrix represent the actual and predicted values respectively. The four most important terminologies in a confusion matrix are True Positive (TP), True Negative (TN), False Positive (FP) or Type-I Error, and False Negative (FN) or Type-II Error. True Positive (TP): The model correctly

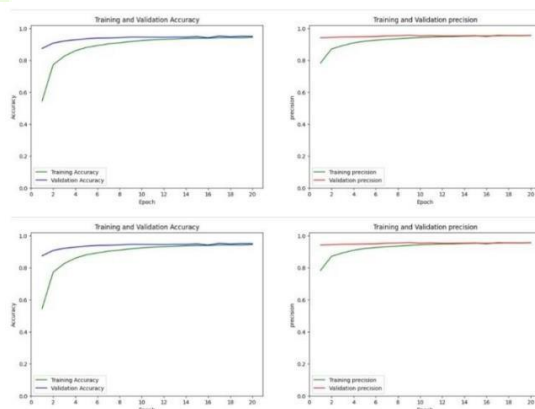
predicted a positive outcome, where the predicted value matches the actual value. True Negative (TN): The model correctly predicted a negative outcome, where the predicted value matches the actual value. False Positive (FP) or Type-I Error: The model predicted positive but it is false, which means the predicted value was falsely predicted. The actual value was negative but the model predicted as positive. False Negative (FN) or Type -II Error: The model predicted negative but it is false, which means the predicted value was falsely predicted. The actual value was positive but the model predicted as negative.

Accuracy: Accuracy refers to the proportion of correct predictions made by our model among all cases. It is determined by dividing the total number of right predictions by the total number of instances in the dataset.

Precision: Precision indicates how many of the instances that were accurately predicted turned out to be positive. It decides whether a model is trustworthy or not. It's beneficial in situations when a False Positive (FP) is more of a concern than a False Negative (FN). The formula for calculating the precision is following: Precision = $TP / TP + FP$ (6)

Recall: Recall indicates how many of the actual positive cases our model was able to correctly anticipate. It is a useful metric in cases where False Negative (FN) trumps False Positive (FP). The formula for calculating the precision is following: Recall = $TP / TP + FN$ (7)

F1 Score: F1 Score also known as F-Score or F-Measure is a harmonic mean of Precision and Recall. F-Score comes handy where it is difficult to compare to models with low precision and high recall or viceversa. The formula for calculating the precision is following: F1 Score = $2 \times \text{Precision} \times \text{Recall} / \text{Precision} + \text{Recall}$ (8)



5. CONCLUSION

The integrated system effectively combines two key components of precision agriculture: crop recommendation and plant leaf disease detection. The crop recommendation module, using Gaussian NB, achieves remarkable accuracy in recommending crops based on environmental factors. The leaf disease detection system, powered by deep learning models like VGG19, DenseNet201, and InceptionV3, provides accurate disease detection results. The ensemble model further improves performance, offering 98.3% accuracy in disease classification. This integrated system empowers farmers with data-driven crop recommendations and early disease diagnosis, enabling improved resource management, enhanced yields, and sustainable agricultural practices. The system's high accuracy in crop recommendation and disease detection underscores its reliability and effectiveness in addressing real-world agricultural challenges.

REFERENCES

- [1] Plant village dataset: <https://www.kaggle.com/datasets/vipooooool/new-plantdiseases-dataset>
- [2] Crop recommendation dataset:
<https://www.kaggle.com/datasets/atharvaingle/croprecommendation-dataset>
- [3] M. H. Bijoy, N. Hasan, M. Biswas, S. Mazumdar, A. Jimenez, F. Ahmed, M. Rasheduzzaman, and S. Momen, "Towards sustainable agriculture: A novel approach for rice leaf disease detection using dCNN and enhanced dataset," IEEE Access, vol. 12, pp. 34174–34182, Feb. 2024, doi: 10.1109/ACCESS.2024.3371511.
- [4] C. Madhurya and E. A. Jubilson, "YR2S: Efficient deep learning technique for detecting and classifying plant leaf diseases," IEEE Access, vol. 12, pp. 3790–3797, Dec. 2023, doi: 10.1109/ACCESS.2023.3343450
- [5] Patel, S., & Dabhi, V. K. (2020). "Hybrid Machine Learning Approach for Crop Recommendation." International Journal of Emerging Trends & Technology in Computer Science, Vol. 9, Issue 2.
- [6] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). "Using Deep Learning for Image-Based Plant Disease Detection." Frontiers in Plant Science, Vol. 7.
- [7] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). "Deep Learning in Agriculture: A Survey." Computers and Electronics in Agriculture, Vol. 147.
- [8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). "Densely Connected Convolutional Networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261 – 2269.
- [9] Jadhav, A., & Gholap, A. (2021). "Crop Recommendation System Using Machine Learning Algorithms." International Journal of Scientific Research in Computer Science Engineering and Information Technology, Vol. 7, Issue 1.
- [10] Ganguly, A., et al., "Ensemble Learning for Plant Leaf Disease Detection: A Novel Approach for Improved Classification Accuracy," VIT Bhopal University, Aug. 17, 2023.
- [11] "Design and Implementation of an Integrated System for Crop Recommendation and Leaf Disease Detection," Project Basepaper.
- [12] Islam, M. M., et al., "DeepCrop: Deep learning-based crop disease prediction with web application," Journal of Agriculture and Food Research, vol. 14, 2023.