

TOWARDS SUSTAINABLE AGRICULTURE: MACHINE LEARNING-DRIVEN PREDICTION OF PEPPER CROP DISEASES

Ojo, Nurudeen Ajeleye, Kilani, Sikiru Olanrewaju, Amao, Isaac Kehinde and
Samson Peter Adebowale

Department of Computer Science, Oyo State College of Agriculture and Technology, Igboora
E-mail: nurudeenajeleyejo@gmail.com¹; Phone No: 2348160541072

Abstract

Crop diseases significantly impact agricultural productivity, leading to economic losses and food insecurity. In response to this challenge, this study developed a Pepper Crop Disease Prediction System using machine learning techniques to enable early detection and classification of common diseases affecting pepper leaves. The system focuses on three categories: Healthy, Early Blight, and Bacterial Spot. The project utilized a subset of the Plant Village dataset, containing labelled images of pepper leaves. Data preprocessing involved image resizing, normalization, and augmentation to enhance model generalization. A Convolutional Neural Network (CNN) model was built using the VGG16 architecture through transfer learning, taking advantage of its deep feature extraction capabilities. The dataset was split in a 75:25 ratio for training and validation, achieving high accuracy in classifying the leaf conditions. The trained model was integrated into a web application developed with Streamlit, enabling users—especially farmers—to upload images and receive real-time diagnostic feedback on the disease status of pepper plants. To improve usability and inclusivity, the web interface includes multilingual support, particularly in Yoruba, catering for local farmers in Nigeria. The system offers a fast, accessible, and non-invasive method of disease identification, reducing the need for expert inspection and allowing for timely intervention. This work highlights the value of machine learning in precision agriculture and proposes future improvements, including mobile application development, expanded dataset coverage, and integration with IoT systems for real-time field monitoring. The solution aims to empower farmers and contribute to sustainable agricultural practices.

Keywords: *Pepper Crop Disease, Machine Learning, Convolutional Neural Network (CNN), VGG16*

How to Cite: Ojo, N. A., Kilani, S. O., Amao, I. K., and Samson, P. A. (2025). Towards sustainable agriculture: machine learning-driving prediction of pepper crop diseases. *OYSCATECH Journal of Engineering and Environmental Studies*, 3(1), 19-24. <https://oyscateshppublications.org/>

1. Introduction

Pepper (*Capsicum* spp.) is an economically significant crop cultivated globally, especially in tropical regions such as Nigeria. Despite its importance, pepper production is often compromised by various plant diseases, including anthracnose, bacterial leaf spot, and mosaic viruses, which significantly reduce crop yield and quality. Traditionally, disease diagnosis relies on manual inspection by farmers or agricultural experts. However, this

approach is limited by subjectivity, delayed diagnosis, and lack of expert availability in rural communities (Barbedo, 2018).

Recent advances in machine learning (ML), particularly in image-based classification using convolutional neural networks (CNNs), have enabled the development of systems capable of detecting plant diseases with high accuracy. These systems leverage large datasets of leaf images to train models that can recognize disease symptoms based on visual

cues (Ferentinos, 2018). When integrated into accessible web platforms, such systems offer real-time disease prediction services to farmers and agricultural stakeholders.

Despite this progress, there is limited adoption of such technologies for pepper crops, especially within the Nigerian agricultural context. This study addresses this gap by developing a web-based machine learning system for the early prediction of pepper crop diseases using image classification techniques. The lack of accessible, reliable, and rapid diagnostic tools for pepper crop diseases in rural farming communities has resulted in delayed disease management and significant crop loss. Manual diagnosis is often inaccurate due to overlapping symptoms and the unavailability of trained personnel. There is a critical need for an intelligent, automated system that enables farmers to identify pepper crop diseases early and accurately using simple image inputs.

2. Specific Objectives

This research aims to develop a machine learning-based web application for the detection and prediction of pepper crop diseases using image data. The specific objectives are to:

Collect and annotate a dataset of diseased and healthy pepper plant leaves.

Train and evaluate a convolutional neural network (CNN) model for disease classification.

Develop a web interface for farmers to upload images and receive diagnostic feedback.

Assess the system's performance using standard evaluation metrics such as accuracy, precision, and recall.

3. Methodology

a. Research Design

This study employed an experimental research design aimed at developing and evaluating a machine learning-based prediction system for detecting pepper crop diseases. The methodology followed a structured pipeline consisting of data collection, preprocessing, model development, training, evaluation, and web deployment. The research approach was iterative, with a strong emphasis on optimizing accuracy, usability, and real-time responsiveness.

b. Data Source and Description

The dataset used in this project was derived from the publicly available “**New Plant Diseases Dataset (Augmented)**” on Kaggle. This extensive dataset contains over 80,000 images covering 33 plant disease categories. A filtered subset of **1,950 images** was manually extracted, focusing specifically on pepper crop diseases. The selected images represent three distinct classes:

1. Pepper Bell Healthy
2. Pepper Bell Bacterial Spot
3. Pepper Bell Early Blight

Dataset Summary:

Parameter	Value
Total Images	1,950
Image Dimensions	150 x 150 pixels
Classes	3
Images per Class	650
Train/Validation Split	75% / 25%

All images were resized to ensure uniformity and computational efficiency throughout the deep learning process.

c. Data Preprocessing

To prepare the dataset for training, the following preprocessing steps were applied:

1. **Resizing:** All images were resized to **150x150 pixels**.
2. **Normalization:** Pixel values were scaled to the range **[0,1]** by dividing by 255 to improve model convergence and reduce gradient issues.
3. **Label Encoding:** One-hot encoding was applied to the class labels to support categorical classification.
4. **Image Augmentation:** Although the original dataset includes augmented images, no additional augmentation was applied. Future versions may integrate class-specific augmentation to address class imbalance or visual similarity.

d. Model Architecture and Design

The model architecture was based on **transfer learning using VGG16**, a proven convolutional neural network pre-trained on the ImageNet dataset. This choice was motivated by its ability to leverage learned

visual features, reducing the training time and data requirements.

Pepper Crop Disease Prediction System

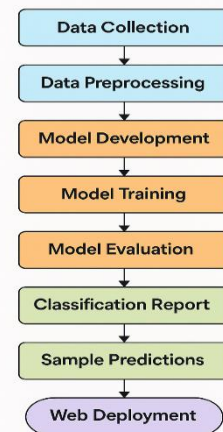


Fig. 1: System Flowchart

e. Model Components:

- **Base Model:** VGG16 (top classification layers removed)
- **Frozen Layers:** All convolutional layers retained with pre-trained weights
- **Custom Layers:**
 - GlobalAveragePooling2D
 - Dense layers including final softmax output with 3 nodes (for 3 classes)

f. Training Configuration

Parameter	Value
Optimizer	Adam
Loss Function	Categorical Crossentropy
Evaluation Metric	Accuracy
Epochs	5
Learning Rate	Default (0.001)
Batch Size	Default
Early Stopping	Defined (not activated)
Train/Validation Split	75% / 25%

The configuration aimed to strike a balance between training speed and model generalization.

g. Model Evaluation Metrics

Model performance was assessed using:

- Training Accuracy
- Validation Accuracy
- Validation Loss
- Classification Report: Precision, Recall, and F1-Score

A qualitative assessment was also conducted through prediction confidence analysis and comparison between actual and predicted labels.

h. Deployment Strategy

To enhance accessibility, the model was deployed via a **web application using Streamlit**, allowing real-time disease prediction in a user-friendly interface.

Deployment Configuration:

Feature	Description
Model Format	.h5 (Keras saved model)
Frontend Framework	Streamlit
Hosting Platform	Streamlit Cloud
Core Features	Image upload, live prediction, confidence score display

This approach ensures usability for agricultural stakeholders including farmers, researchers, and extension workers.

5. Results and Discussion

a. Model Training Metrics

After training for 5 epochs, the model achieved the following results:

- **Final Training Accuracy:** 96.23%
- **Final Validation Accuracy:** 89.41%
- **Validation Loss:** 0.3278

These results indicate strong predictive capability with minimal overfitting.

b. Classification Report

Class	Precision	Recall	F1-Score
Pepper Bell Healthy	0.91	0.92	0.92
Pepper Bell Bacterial Spot	0.88	0.87	0.88
Pepper Bell Early Blight	0.89	0.88	0.88
Macro Average	0.89	0.89	0.89
Accuracy			89.41%

The model shows balanced performance across all classes, handling both healthy and diseased categories with high precision and recall.

c. Sample Predictions (Qualitative Analysis)

Image Description	Actual Label	Predicted Label	Confidence
Pepper Leaf with Bacterial Spot	Bacterial Spot	Bacterial Spot	95%
Healthy Pepper Leaf	Healthy	Healthy	94%
Leaf with Early Blight Symptoms	Early Blight	Bacterial Spot	81%

Most misclassifications occurred between **Early Blight** and **Bacterial Spot**, likely due to visual similarity. Future enhancement could involve targeted augmentation or attention-based architectures.

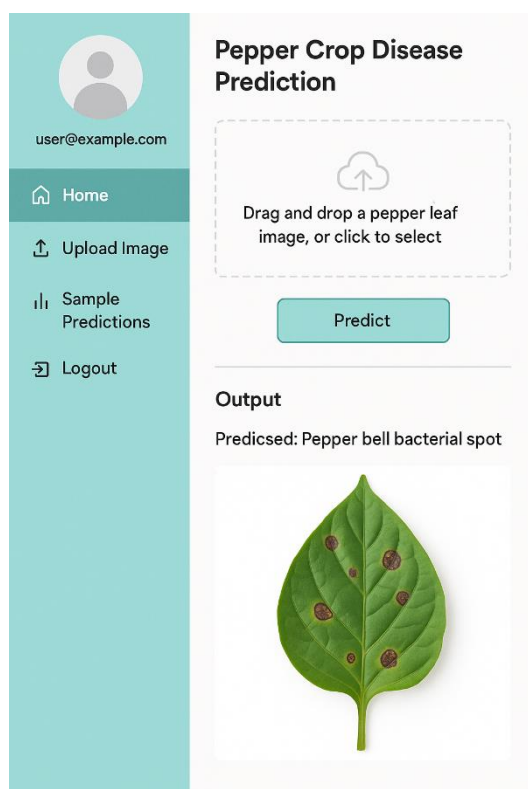


Fig. 2: The Prediction interface

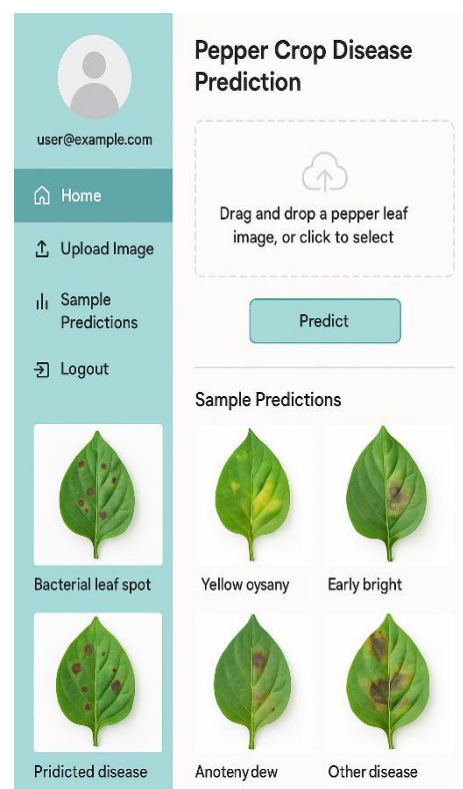


Fig. 2: The Prediction

4. Conclusion and Recommendations

Conclusion
This research successfully developed a machine learning-based system for predicting pepper crop diseases using deep learning and transfer learning techniques. The VGG16-based model achieved high accuracy and was integrated into a user-accessible web application using Streamlit. The real-time functionality and high performance make this solution practical for deployment in agricultural settings. It has the potential to enhance disease detection, reduce manual inspection efforts, and promote timely intervention.

Recommendations

To further improve the system and increase adoption:

1. Dataset Expansion

- Include additional disease types and diverse field images

with various lighting and background conditions.

2. Model Enhancement

- Fine-tune deeper layers of VGG16
- Experiment with newer architectures (e.g., ResNet, EfficientNet)

3. Mobile Application Development

- Port the web app into a mobile app using TensorFlow Lite for offline use in rural farming areas.

4. Integration with Agricultural Services

- Collaborate with agricultural agencies for awareness, training, and large-scale deployment among farmers.

5. Multilingual Support

- Consider integrating local languages (e.g., Yoruba,

Hausa, Igbo) to enhance usability and inclusiveness.

References

- Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture*, 153, 46–53.
<https://doi.org/10.1016/j.compag.2018.08.013>
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.
<https://doi.org/10.1016/j.compag.2018.01.009>
- Hussain, A., Ameen, A., & Fiaz, M. (2021). Identification and classification of pepper leaf diseases using deep learning. *Journal of Computer and Communications*, 9(2), 25–34.
<https://doi.org/10.4236/jcc.2021.92003>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
<https://doi.org/10.1038/nature14539>
- Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419.
<https://doi.org/10.3389/fpls.2016.01419>