International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 14 Issue 7, July-2025

Maize Leaf Disease Identification for Sustainable Crop Productivity

K. Deepika[1], Dr. K. Manideep[2]

[1]M-Tech with Specialization of Computer Science and Engineering, Bapatla Engineering College, Bapatla
[2]Associate Professor, Department of Computer Science & Engineering (AI & ML), Bapatla Engineering College, Bapatla

ABSTRACT

Maize is one of the world's most important agricultural crops in the world which is affected by various pathogenetic diseases. These diseases lead to low productivity and huge loss for the farmers. For this reason, detection of these diseases in early stage by recognizing their symptomatic patterns will be beneficial for farmers. For this reason, we have created a dataset with 29065 maize disease images where 80% of samples were used for training and achieved accurate 99.02%. This study explores the use of convolutional neural networks (CNNs) to automatically detect and classify common maize leaf diseases, such as Northern Leaf Blight, Rust, and Gray Leaf Spot.

Keywords: Deep Learning, Identification, Maize leaves disease classification, CNN architecture.

INTRODUCTION

Maize (Zea mays) is one of the most widely cultivated cereal crops, serving as a staple food for millions of people worldwide. However, maize productivity is often threatened by various leaf diseases, such as Northern Corn Leaf Blight (NCLB), Common Rust, and Gray Leaf Spot.

Plant diseases present a significant menace to both food security and the global economy. Not only maize, but also major crops like wheat, rice, and soybeans are susceptible to various types of diseases caused by pathogens. These diseases adversely affect crops, farmers, and the agricultural industry. Farmers face decreased yields, lower quality crops, and increased costs for pesticides and fungicides.

RELATED WORK

Deep learning algorithms have excelled in diagnosing plant diseases with remarkable accuracy. They offer a more effective and precise alternative to traditional machine learning methods. CNNs are widely used for image classification tasks and have achieved state-of-the-art performance.

These models outperform traditional methods in plant disease classification. CNNs demonstrated higher accuracy than earlier models in both original and modified image simulations.

MATERIAL AND METHODS

Datasets:
Deep learning models excel at learning from large datasets but depend heavily on data quality and quantity. Effective datasets must be well-curated and representative of real-world scenarios. This study uses maize leaf images from PlantVillage (4 classes), PlantDOC (3 classes), and CD&S (3 classes). Classes with identical diseases were merged to form a unified dataset. The combined dataset improves model training and generalization.

Plant Village Dataset
This study uses a subset of the PlantVillage dataset, containing 3,839 maize leaf images from four categories, including Common Rust, Healthy, and Northern Leaf Blight.

International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 14 Issue 7, July-2025

The original dataset comprises 54,303 leaf images classified into 38 disease types, both healthy and diseased. For balanced evaluation, data is split into 70% training, 15% validation, and 15% testing. This ensures model generalization and objective performance measurement.

PlantDoc Dataset

Another significant part of our larger dataset is the publicly available PlantDoc dataset, which contains diseases of various plants. PlantDoc is a dataset designed for visual plant disease identification. Some images of maize leaf diseases from the PlantDoc dataset are

Figure: 1. Images of healthy and diseased maize leaves from Plant Village dataset.

Table 1: Statistical data of maize classes in the Plant Village dataset.

Class Names	Total (% 100)	Train (% 70)	Validation (% 15)	Test (% 15)
Gray Leaf Spot	513	359	77	77
Common Rust	1186	830	178	178
Northern Leaf Blight	978	684	147	147
Healthy	1162	812	175	175

PlantDoc dataset has 2,598 data points overall spanning 13 plant species and as many as 17 disease classes, and it required about 300 person hours to add descriptions to photographs that were downloaded from the internet. To generate our new dataset, we combined the classes from the PlantDoc dataset with the equivalent classes from other datasets. For each class, the data was split into three categories: training (70 %), validation (15 %), and testing (15 %), which is depicted in Table 2.

New Dataset:

We created a new dataset by merging identical classes from PlantVillage, PlantDoc, and CD&S. Each dataset was split into training, validation, and testing sets in a 70:15:15 ratio. The merged dataset includes classes like Cercospora Leaf Spot, Common Rust, Healthy, and Northern Leaf Blight from PlantVillage. It also includes Maize Leaf Blight, Gray Leaf Spot, and Rust Leaf from PlantDoc. Gray Leaf Spot and Northern Leaf Blight from CD&S were added for greater diversity.

Figure: 2. Images of diseased maize leaves from PlantDoc dataset

Table 2: Statistical data of maize classes in the PlantVillage dataset

Class Names	Total (% 100)	Train (% 70)	Validation (% 15)	Test (% 15)
Maize Leaf Blight	191	133	29	29
Maize Gray Leaf Spot	68	48	10	10
Maize Rust Leaf	116	82	17	17

Deep Learning Architectures:
The field of artificial intelligence has been revolutionized by deep learning, which has made it possible to create sophisticated algorithms capable of learning from large volumes of data. One of the most significant applications of deep learning is computer vision, such as facial recognition, medical image analysis, and self-driving cars.

Table 3 Details of the newly constructed dataset

Class Names	Total (% 100)	Train (% 70)	Validation (% 15)	Test (% 15)
Gray Leaf Spot	1104	774	165	165
Leaf Blight	1666	1164	251	251
Rust Leaf	1302	912	195	195
Healthy	1162	813	175	175

Proposed Method
Timely and accurate detection of plant leaf diseases is vital for efficient agricultural production. Deep learning algorithms, particularly CNNs, are widely used for disease classification and detection. CNNs are designed for processing grid-like data such as images. They can be implemented using frameworks like TensorFlow, PyTorch, or Keras. Preprocessing includes image resizing, augmentation, and dataset splitting. Transfer learning with models like ResNet, VGG16, or EfficientNet boosts performance.

Data Collection and Preprocessing

- Dataset: Gather maize leaf images from open datasets or field sources.
- Data Augmentation: Apply techniques like rotation, scaling, contrast adjustment, and noise addition to improve model generalization.
- Image Preprocessing: Resize images to a standard size, normalize pixel values, convert to grayscale or RGB as needed

RESULTS AND DISCUSSIONS

Performance Metrics
Performance metrics are essential for assessing deep learning models. Accuracy measures the overall correctness of predictions. Precision and recall offer insights into true positives and accurate positive predictions. Recall shows how many

actual positive examples were correctly identified. The F1 score balances precision and recall by taking their harmonic meaning. These metrics are particularly important for unbalanced datasets.

Accuracy= Number of correct prediction/ Number of total Predictions

Precision= TruePositive/ TruePositive +FalsePositive

Recall= Sensitivity= TruePositive/ TruePositive +FalseNegative

F1=2*Precision*Recall / Precision + Recall

Training Procedure

To increase the deep learning model's training accuracy and speed, various methods and variables can be used. Two of the best methods available are transfer learning and data augmentation. Other parameters like warmup epochs and warmup learning rate can be fine-tuned to gradually increase the learning rate during the initial training epochs to avoid divergence.

Results

This section discusses the experimental results, performance measurements, and confusion matrix of CNN and vision transformer models. All results were evaluated on separate test datasets, pre-separated from 15% of newly generated data. Typically, datasets are split into training and validation sets, with performance measured on the validation set. However, this approach doesn't fully demonstrate the models' generalization capabilities.

Results for the CNN models

CNN architectures are one of the most used architectures in image-based artificial intelligence applications in deep learning. Many CNN architectures have been presented in the literature and have become popular architectures in various applications such as medicine, agriculture, defense, and autonomous vehicles. In this study, a detailed analysis is presented using the most popular CNN models found in the literature to train and evaluate a newly created larger dataset of maize leaf diseases.

Fig. 3. A batch of maize leaf image from a new merged dataset used during the training process.

The misclassification rate measures inaccurate predictions, calculated as (FP + FN) / total instances). "Gray Leaf Spot" consistently has the highest misclassification rate among the evaluated models. The Healthy class achieves perfect scores in Precision, Recall, and F1-score (1.0).

Other classes also perform well, with high Precision (0.9763 for Gray Leaf Spot, 1.0 for Leaf Blight, 0.9898 for Rust Leaf). Recall scores are nearly perfect for all classes, except for Leaf Blight.

Table 4: Class-wise report of the Proposed Model

Metrics/Classes	Gray leaf_spot	Leaf Blight	Rust Leaf	Healthy
Precision	0.9763	1.0	1.0	0.9898
Recall	1.0	1.0	0.9801	0.9949
F1-score	0.9880	1.0	0.9899	0.9923
No-Images	165	175	251	195

Limitations and Future Research

The model achieves 99.24% accuracy, potentially transforming crop management. It can reduce disease-related losses, decrease herbicide use, and boost farmers' financial returns. CNNs are deep learning models designed for processing structured data like images. They excel in tasks such as image classification, object detection, and segmentation. CNNs are effective due to their ability to capture spatial hierarchies in data.

Confusion Matrix

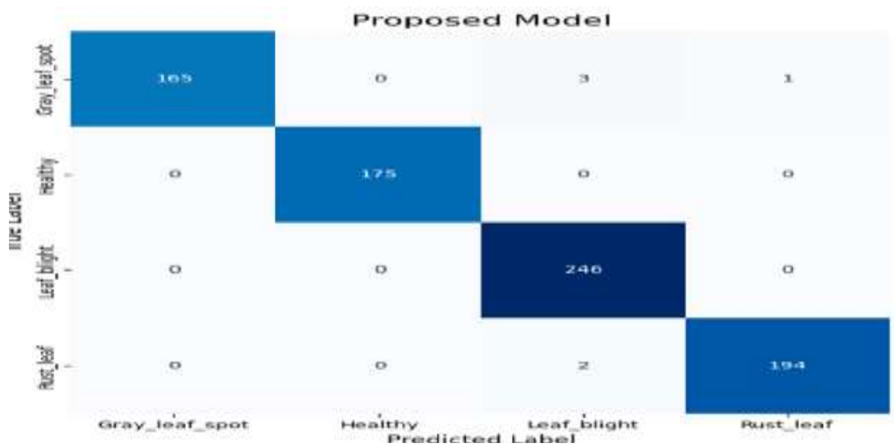

Figure: 4. Confusion Matrix of the Proposed Model

CONCLUSION

This study introduces a deep learning-based CAD system for early detection of maize leaf diseases. It utilizes vision transformers to achieve superior performance over CNN and other transformer models. The Proposed Model attains 99.24% accuracy, indicating strong predictive power. It also excels in precision, recall, and F1 score metrics. Efficient GPU memory use and fast inference enhance practicality. These optimizations contribute to its overall improved performance.

REFERENCES

[1]. Ahmad, A., Saraswat, D., Gamal, A. El, & Johal, G. (2021). *CD&S Dataset: Handheld Imagery Dataset Acquired Under Field Conditions for Corn Disease Identification and Severity Estimation*.http://arxiv.org/abs/2110.12084.

[2]. Alqahtani, Y., Nawaz, M., Nazir, T., Javed, A., Jeribi, F., & Tahir, A. (2023). An improved deep learning approach for localization and recognition of plant leaf diseases.*Expert Systems with Applications*, *230*.https://doi.org/10.1016/j.eswa.2023.120717

[3]. Ashwini, C., & Sellam, V. (2023). EOS-3D-DCNN: Ebola optimization search-based 3D-dense convolutional neural network for corn leaf disease prediction. *Neural Computing and Applications*, *35*(15), 11125–11139. https://doi.org/10.1007/s00521- 023-08289-3

[4]. Cui, S., Su, Y. L., Duan, K., & Liu, Y. (2023). Maize leaf disease classification using CBAM and lightweight Autoencoder network.*Journal of Ambient Intelligence and Humanized Computing*, *14*(6), 7297–7307.https://doi.org/10.1007/s12652-022-04438-z

[5]. Divyanth, L. G., Ahmad, A., & Saraswat, D. (2023). A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery. *Smart Agricultural Technology*, *3*, Article 100108.https://doi.org/10.1016/j.atech.2022.100108