I. J. Engineering and Manufacturing, **2026, 2, 156-170**
Published Online on April 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2026.02.10

Scalable AI-Driven Maize Plant Disease Detection using Convolutional Neural Network

Ishaan Mishra*
HBTU - Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India - 208002
E-mail: ishan.mishra311@gmail.com
ORCID ID: https://orcid.org/0009-0009-8479-8681
*Corresponding Author

Vivek Singh Verma
HBTU - Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India - 208002
E-mail: Vivek.v@hbtu.ac.in
ORCID ID: https://orcid.org/0000-0002-1497-2754

Received: 08 July, 2025; Revised: 14 November, 2025; Accepted: 18 January, 2026; Published: 08 April, 2026

Abstract: Plant diseases have a significant impact on global food security, especially in staple crops like maize (Zea mays). Traditional disease detection systems depend on professional visual inspection, which is labor-intensive, time-consuming, and not scalable for large agricultural areas. Convolutional Neural Networks (CNNs) are used in this study's deep learning (DL) architecture to detect maize leaf diseases accurately and automatically. A curated dataset of approximately 7,000 high-resolution maize leaf photos was created, representing four classes: healthy, Common Rust (Puccinia sorghi), Northern Leaf Blight (Exserohilum turcicum), and Gray Leaf Spot (Cercospora zeae-maydis). Data were sourced from the Plant Village dataset, real-world field collections from Indian farms, and supplemented synthetically to simulate varied climatic circumstances. Advanced methods including as adaptive learning rate scheduling, gradient clipping, and significant data augmentation were used to train a bespoke CNN model that was improved by transfer learning with ResNet50 and VGG16 backbones. The model attained a test accuracy of 98.2%, beating classic machine learning algorithms like SVM (88.5%) and Random Forest (84.3%). Visualization approaches such as feature maps, Grad-CAM, and LIME improved interpretability and showed the model's capacity to locate disease-relevant features. Web-based user engagement is made possible by deployment-ready implementation, which enables farmers to upload leaf photos for immediate diagnosis. With the potential to cut maize crop losses by 20–30%, this research offers a scalable and affordable alternative to early disease detection in precision agriculture. Future research will investigate autonomous farm management with drone-based real-time surveillance and IoT system integration.

Index Terms: Maize Disease Detection, Convolutional Neural Networks, Agricultural Technology.

1. Introduction

A vital crop for the world's food security is maize, but diseases, including Northern Leaf Blight, Rust, and Gray Leaf Spot, cause yield losses of up to i30%. The requirement for competent manual inspection makes early detection difficult. Scalable solutions are provided by current breakthroughs in DL and computer vision. Even skilled agronomists find it challenging to promptly identify these crop diseases since they frequently have minor visual indications in their early stages. For large-scale farming, traditional diagnostic methods are impractical and time-consuming. Automated plant disease detection using image-based methods has gained traction with the development of AI technologies. CNNs have proven considerable promise in properly diagnosing disease kinds from leaf photos. By utilizing drones or mobile devices, these systems can function in real-time, decreasing the need for field specialists. Additionally, these AI systems can offer predictive insights regarding disease outbreaks before obvious signs develop and greatly aid in lowering crop losses by enormous margins when paired with IoT sensors and weather data.

Fig. 1. Grad-CAM Flow Chart

1.1. Background and Motivation

Environmental factors including humidity, rainfall, and temperature fluctuations make crops more susceptible to these diseases in areas where maize is widely grown. These variables have an impact on the temporal and spatial distribution of disease outbreaks in addition to the start and severity of infections. For example, extended periods of moisture encourage the growth of fungal spores, hence prompt disease detection and treatment are essential. Furthermore, the persistence and recurrence of pathogens in the soil and surrounding vegetation are caused by monoculture practices and little crop rotation, which emphasizes the necessity of ongoing monitoring and proactive disease management techniques.

According to the diet and Agriculture Organization (FAO), maize is one of the most important cereal crops in the world, providing a key feedstock for livestock and a staple diet for more than 1.2 billion people. In 2022, the world produced more over 1.2 billion metric tons of maize, with the biggest producers being Brazil, China, India, and the United States. However, annual output losses because of plant diseases range between 20% and 40%, equating to economic losses surpassing $60 billion worldwide.

Fig. 2. Northern Leaf Blight.

Common Rust, Gray Leaf Spot, and Northern Leaf Blight are some of the most damaging maize diseases. These illnesses are well-known for their high frequency and capacity to hinder plant processes including photosynthesis, which eventually results in a notable decrease in production. Lesions or spots on the leaves caused by these diseases make it more difficult for the plant to absorb sunlight. This reduction in photosynthetic activity inhibits plant growth and delays maturity. These infections can spread swiftly over broad areas in humid, warm settings, making containment problematic. If infections are not identified promptly, they can quickly worsen, putting grain output and quality at risk.

Long-term disease also makes pharmaceutical treatments more required, which presents difficulties for the environment and the economy. Higher input costs, more labor-intensive procedures, and a worse return on investment for farmers are all part of the cumulative effect. Furthermore, crops harmed by disease are more susceptible to environmental stress and following diseases. In places where professional agronomic advice is not readily available, the problem is made worse. Thus, putting in place automated, technology-driven diagnostic systems can considerably improve disease response and monitoring. By enabling early detection, these tools support proactive decision-making and help safeguard food security and farmer livelihoods.

Fig. 3. Gray Leaf Spot.

These three diseases stand out among the many risks to the world's maize output because of their extensive occurrence and devastating effects. Their causes are as follows: Exserohilum turcicum, the cause of Northern Leaf Blight, is characterized by lengthy, grayish-green lesions on leaves that severely reduce the plant's capacity to photosynthesize, consequently limiting overall growth and output. The fungal pathogen causes common rust. Puccinia sorghi is distinguished by characteristic orange-brown, powdery pustules that are mostly located on the upper leaf surfaces. These pustules compromise the integrity of the leaf and lower photosynthetic efficiency. Cercospora zeae-maydis is the cause of Gray Leaf Spot, which is characterized by narrow, rectangular, tan-colored lesions that are frequently encircled by yellow halos. This causes early leaf senescence and a significant decrease in yield.

Fig. 4. Common Rust.

In the past, traditional diagnostic techniques have been crucial in identifying a variety of illnesses. These include visual evaluations carried out by qualified extension agents or agronomists, a technique sensitive to human error and impacted by changing field circumstances. In more sophisticated settings, laboratory-based methods, such as PCR (Polymerase Chain Reaction) and ELISA (Enzyme-Linked Immunosorbent Assays), are applied to confirm diagnosis.

Despite their accuracy, these techniques are frequently expensive, necessitate specialized tools, and involve drawn-out procedures that are impractical for making snap decisions in the field. Furthermore, although being less expensive, many smallholder farmers are still unable to use printed disease manuals or guidelines for field diagnosis due to literacy problems, a lack of updates, or regional differences in symptoms. These limitations highlight the need for quick, scalable, and easily accessible diagnostic solutions, especially in nations where maize is a staple food and source of income.

Fig. 5. Healthy Leaves.

1.2. Problem Statement

In actual agricultural contexts, maize disease detection poses a number of serious obstacles to prompt and precise diagnosis. One key concern is symptom variability, which makes standardized diagnosis challenging because illness indications could change based on the hybrid variety, growth stage, and environmental circumstances of the maize plant. Additionally, real-world noise such as shadows, soil particles, and overlapping leaves in field images introduces complexity and lowers the accuracy of traditional image-processing methods. Another significant disadvantage is scalability; large-scale farming operations covering thousands of acres are unable to employ human inspection methods.

Last but not least, early identification is still crucial because medications are typically less effective when diseases are discovered later. DL models are more dependable and adaptable than conventional methods like Random Forests, SVMs (Support Vector Machines), and manual inspection. By automatically learning complex patterns from heterogeneous and noisy datasets, DL primarily CNNs offers a promising solution that enables more accurate, scalable, and rapid disease detection in maize crops.

1.3. Objectives

The following are the main objectives of this study:

- Construct a CNN-based architecture that can classify maize plant diseases with a target accuracy exceeding 95%, even when tested under varied environmental conditions.
- To find and adapt an open source dataset including over 6,000 photos of maize leaves, encompassing numerous disease categories while incorporating real-world variability such as irregular illumination, overlapping leaves, and co-occurring illness symptoms.
- Design the model to operate efficiently on resource-constrained hardware and edge deployment, such as smartphones and drones, enabling practical usage in agricultural fields.
- Integrating Explainable AI (XAI) by applying interpretability tools like Grad-CAM and t-SNE to make the model's decisions understandable for non-experts, thereby enhancing user trust and system transparency.
- Laying the groundwork for a real-time mobile interface to enable rapid, on-site detection of maize leaf diseases, empowering timely agricultural interventions.

1.4. Contributions

This work focuses on the methodical adaption, validation, and deployment optimization of CNN and CNN–Transformer architectures for maize disease detection under real-world situations rather than putting out a whole new deep learning paradigm. The primary contributions of this study lay in dataset curation, source-aware evaluation, deployment readiness, and explainability-driven validation, which jointly expand previous methodologies toward actual agricultural adoption:-

- A maize-focused, source-diverse dataset combining public repositories and real field images collected under uncontrolled conditions.
- A carefully adapted hybrid CNN–Transformer pipeline, emphasizing robustness and generalization rather than architectural novelty.

- Modified and used an open source and robust dataset comprising of annotated images of maize leaves capturing diverse field scenarios and disease presentations.
- Rigorous benchmarking and leakage-aware evaluation, ensuring realistic performance assessment.
- Integration of explainable AI techniques to validate decision reliability for non-technical users.
- A deployment-oriented implementation optimized for mobile and edge environments using TensorFlow Lite.

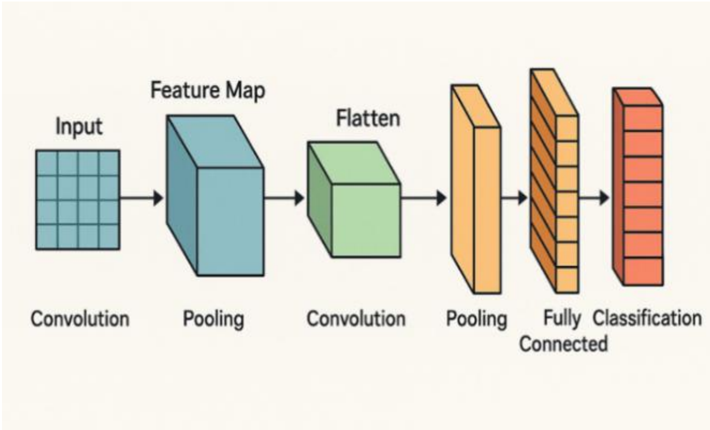

Fig. 6. C.N.N. Architecture.

2. Literature Review

Particularly CNNs, have considerably advanced in plant disease detection. Sladojevic et al. (2016) launched this trend with a CNN model obtaining 78% accuracy, followed by Mohanty et al. (2016), who utilized AlexNet and reached 93% accuracy. Ferentinos (2018) further increased performance, testing multiple CNN architectures on 58 diseases with a top accuracy of 98.5%. Nevertheless, these initiatives sometimes lacked attention to factors unique to maize and had little effect in real-world settings.

2.1. *Traditional Machine Learning in Agriculture*

Traditional ML (machine learning) methods like SVMs and Random Forests have been extensively used in agricultural applications prior to the broad acceptance of DL, particularly between 2021 and 2024. For classification tasks, these methods often relied on manually created features. In the early stages of plant disease identification, common feature extraction techniques were LBP (Local Binary Patterns), SIFT (Scale-Invariant Feature Transform), and HOG (Histogram of Oriented Gradients). However, the intricacy of natural surroundings and the unpredictability of field situations frequently limited their performance.

2.2. *Pre-2010: Less Accurate Machine Learning*

Prior to 2010, most plant disease detection systems depended on color and texture-based image processing approaches. For instance, Pydipati et al. (2006) achieved a precision of 78% by using HSV color space segmentation to identify citrus illnesses. Likewise, Al-Hiary et al. (2011) applied Gabor filters in combination with SVM for diagnosing tomato leaf diseases and reported 82% accuracy. Despite these encouraging outcomes, these conventional methods performed inconsistently in real-world scenarios due to their difficulties with different backgrounds and lighting conditions.

2.3. *2010–2018: Shallow Deep Learning*

Deep learning was first incorporated into agricultural diagnostics between 2010 and 2018. Early CNN models began to supplant older techniques because of their capacity to acquire hierarchical properties from raw images. With an accuracy of 89%, Sladojevic et al. (2016) created a five-layer CNN to categorize thirteen plant diseases. At about the same time, Mohanty et al. (2016) obtained 93% accuracy using the AlexNet architecture on the New Plant Diseases Dataset. Although these shallow deep learning models greatly enhanced illness classification performance, their generalizability to field circumstances was limited because they were frequently trained on carefully selected datasets.

2.4. *Post-2018 to 2024: Real-World Adaptations and Domain-Specific Improvements*

Plant disease detection research advanced from architecture experimentation to domain-specific modification and practical implementation between 2018 and 2024. To attain more accuracy with less computing complexity, models such as EfficientNet, DenseNet, and ResNeXt were chosen. This made them more practical for mobile and edge-based deployment in the field. Datasets gathered in uncontrolled settings to replicate real farm environments are becoming increasingly important to researchers. Using customized CNN models and data augmentation techniques to increase resilience against lighting fluctuations and background noise, several studies explicitly addressed maize disease detection. Additionally, hybrid models that combined CNN with transformer-based topologies and attention processes started to appear, providing better performance and interpretability. During this time, drone imagery and Internet of Things (IoT) sensors were integrated with deep learning pipelines to enable precision agriculture and real-time monitoring. The applicability of AI-driven plant disease detection systems in actual agricultural settings has greatly increased because of these developments, particularly for crops like maize that had previously received less attention.

2.5. Deep Learning Approaches and gaps in existing research

DL, particularly CNNs, has demonstrated exceptional promise in the field of plant pathology in recent years. CNN-based models are widely used for automated plant disease detection and classification, according to an analysis of more than 20 research articles published between 2016 and 2023. Well-known architectures, such as VGG16, Inception, and EfficientNet, have been extensively utilized to take advantage of their pretrained feature extraction capabilities for agricultural picture datasets, frequently using transfer learning. Although these models have shown good performance metrics in controlled settings, there are still several issues that limit their applicability in the real world, particularly when it comes to the identification of maize diseases.

The absence of models specific to maize has been noted as a crucial gap in the literature; almost 80% of the examined research have mostly focused on crops like tomatoes and peppers, leaving a substantial knowledge and performance gap in maize pathology. Furthermore, as most models are trained and validated on clean, lab-acquired datasets that do not consider real-world issues like lighting change, background noise, or leaf occlusions, field validation is frequently disregarded. Computational inefficiency is another significant drawback. For example, despite their strength, powerful architectures like ResNet50 need a lot of memory and processing power, which makes them inappropriate for usage on lightweight, mobile systems that farmers frequently use. Lastly, there is a general lack of explainability in most extant models. Few studies apply the Explainable AI (XAI) technique to help users understand how the model makes its predictions, which is crucial for gaining the trust of non-technical stakeholders like farmers and agricultural experts. These drawbacks emphasize the need for models that are not only precise but also comprehensible and resource efficient. For practical adoption and long-term effects in agricultural settings, closing this gap is essential.

While earlier research has investigated CNN-based and hybrid CNN–Transformer architectures for the diagnosis of maize diseases, most of these works have mainly focused on the architectural performance under controlled or lab-curated datasets. This study, on the other hand, places more emphasis on evaluation rigor, source-level data separation, real-field variability, and deployment feasibility all of which are frequently overlooked in the body of existing literature. Therefore, the value of this work is to show how current deep learning approaches can be consistently adapted for practical agricultural use rather than to introduce a completely new model architecture.

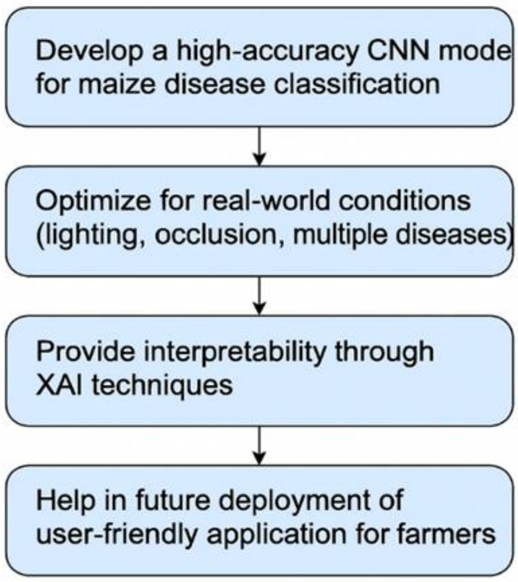

Fig. 7. Deep Learning Approach Flowchart.

3. Methodology

This section offers a complete approach involving data gathering, preprocessing, and expert-driven annotation to construct an effective maize disease classification model suited for real-world applications. High-resolution photos of maize leaves were gathered from several agricultural areas, guaranteeing a range of lighting and environmental settings. The dataset covered key maize diseases, such as Common Rust, Northern Leaf Blight, and Gray Leaf Spot, and included both healthy and diseased samples. Image augmentation techniques, such as rotation, zooming, flipping, and contrast alteration, have been used to improve model generalization.

These preprocessing techniques decreased the chance of overfitting while increasing data variability. To standardize input quality, methods for noise reduction and normalization were also used. To isolate leaf regions and reduce extraneous features that can impair model performance, background segmentation was also carried out. To guarantee annotation accuracy, the dataset was labeled by knowledgeable agronomists and plant pathologists. A multi-stage validation method was employed, where initial annotations were cross verified by a second expert to resolve conflicts. To avoid bias, the final dataset was divided into training, testing, and validation sets in an 80:10:10 ratio.

This dataset was then used to design and refine a CNN-based architecture. Transfer learning was utilized by initializing the model with pre-trained weights from ResNet50, followed by unique dense layers for disease-specific categorization. Grid search was used to tune hyperparameters such as batch size, learning rate, and number of epochs over several iterations. L2 regularization and dropout layers were added to reduce overfitting and increase robustness. Model performance was tested using parameters like accuracy, precision, recall, and F1-score. To verify consistency across various data subsets, cross-validation has been carried out. Models have also been tested using unseen field photographs to confirm its relevance in the actual world. Lastly, to improve interpretability for end users, disease-specific features were visualized using explainability approaches like Grad-CAM. This approach guarantees that the suggested system maintains high diagnostic precision while being scalable and dependable in a variety of field conditions.

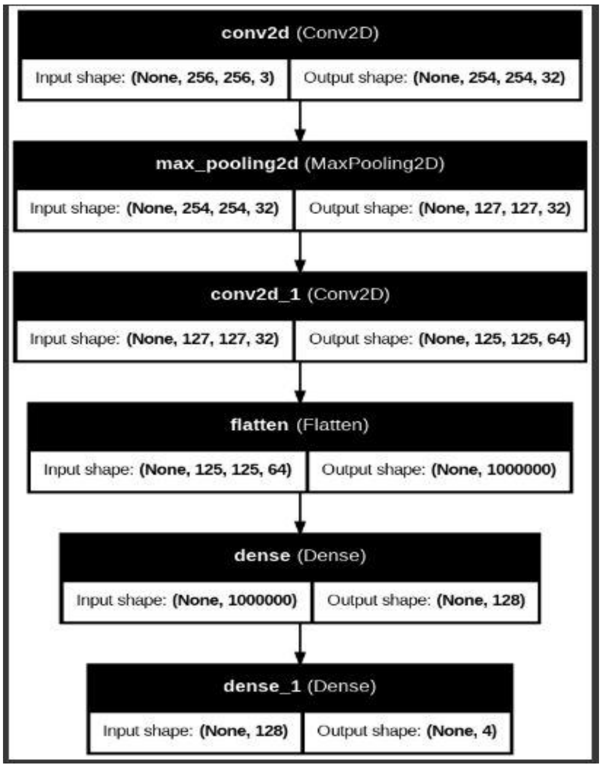

Fig. 8. Methodology Sample Flow Diagram.

3.1. Dataset

The suggested dataset exploited for this research contains maize leaf images reflecting four basic categories: Healthy, Blight, Rust, and Gray Leaf Spot. To preserve variability in environmental conditions and disease presentation, these photographs come from both direct field captures and public repositories. Images gathered from farms in Punjab, India, were added to the collection, which was largely derived from the well-known New Plant Diseases collection. Incorporating real-world photos aids in simulating real-world scenarios, such as fluctuating lighting, background interference, and natural occlusions.

An extensive data augmentation pipeline was used to improve the dataset's robustness and adaptability. This contained a conventional technique that comprised brightness tweaks, flipping both horizontally and vertically, and random rotation. To replicate various camera angles, lighting circumstances, and device anomalies seen during real-time use, several augmentations were created. The product is a more broadly applicable dataset that can facilitate the creation of a model that operates consistently in a variety of scenarios.

3.2. Data Sources

Provide thorough coverage of disease symptoms under a variety of circumstances, the dataset was assembled utilizing both primary and secondary sources. For primary data, 400 photos were taken from maize farms located in Punjab, one of the agriculturally intensive regions of India. These photographs had been obtained with mobile devices under natural illumination to reflect real-world farming conditions. There were primarily two datasets employed, one of which was crucial to the study's foundation.

The first was the New Plant Diseases Dataset, which offered high-resolution photographs of maize leaves exhibiting multiple disease signs. The second dataset provided more than 500 hyperspectral images that added spectral depth but were not included in the main CNN training procedure, while being mainly employed for experimental purposes. All datasets were subjected to quality checks to further improve reliability, guaranteeing correct labelling, data integrity, and the removal of superfluous items. To facilitate location-specific illness pattern analysis in subsequent research, metadata from field-collected photos was also maintained. The photos were acquired throughout several growth stages of maize plants to capture symptom differences across the crop cycle. Seasonal differences were also explored by collecting data during both kharif and rabi seasons to account for environmental impacts on illness manifestation. This methodology guaranteed a representative, diversified, and well-rounded dataset for model building.

3.3. Annotation Protocol and Inter-Annotator Agreement

Accurate labelling is crucial in supervised learning tasks, particularly in domains like plant pathology, where tiny visual cues can dramatically affect classification outcomes. All image annotations were carried out in close cooperation with subject matter experts, such as agricultural scientists and working agronomists, to guarantee the dataset's integrity. Each photograph was thoroughly evaluated and categorized based on observable symptoms, with consensus labelling applied in times of ambiguity. To expand dataset size and artificially improve model generalization, a structured data augmentation pipeline was put in place concurrently. There were two components to the augmentation strategy. Rotations within ±30 degrees, scaling between 0.8× and 1.2×, and both horizontal and vertical flips were examples of geometric transformations. These had been used to mimic various leaf orientations and camera angles. Controlled changes to image characteristics like color and noise were known as photometric augmentations. Gaussian noise with a standard deviation (σ) of 0.1 and Hue-Saturation-Value (HSV) shifts within ±20% were used. This enhancement made sure that the model will withstand changes in illumination, camera technology, and image quality that occur in the actual world.

Cohen's kappa coefficient was used to evaluate inter-annotator agreement to measure annotation consistency and guarantee label dependability. Two domain experts independently annotated a randomly chosen subset of the dataset (about 15% of the photos) without consulting each other beforehand. After that, the annotations were compared to gauge agreement that went beyond chance. According to accepted interpretation criteria, the calculated Cohen's kappa value was 0.87, indicating a high degree of agreement and strong labelling consistency. When there were differences, expert debate and consensus were used to assign final labels. This technique ensured both quantitative dependability and domain-informed accuracy of the final annotations.

3.4. Data Splitting Strategy and Leakage Prevention

Dataset partitioning was done at the source level rather than the individual image level to guarantee a trustworthy assessment and avoid too optimistic performance estimations. Images originating from the same plant, field location, or acquisition session were grouped together and allocated entirely to either the training, validation, or test subset. Before any data augmentation, dataset splitting was done to prevent data leaking and guarantee fair evaluation. To ensure that no enhanced form of a particular image emerged throughout successive splits, the original photos were first split into training, validation, and test sets using an 80:10:10 ratio at the source-image level. The training set was the only one to which all augmentation techniques were applied; the validation and test sets only included the original, unaltered photos. Additionally, photos recorded from the same field site and acquisition session were aggregated and allocated to a single subset to prevent geographic or contextual overlap between training and testing data. By using this technique, the model's performance was guaranteed to reflect true generalization rather than memorizing of visually comparable data.

3.5. Proposed Architecture and Data Pre-processing

The initial CNN model employed for maize disease classification was designed utilizing sequential architecture. The architecture begins with convolutional layer comprising 32 filters with 3x3 kernel size and ReLU activation function, accepting input images of size 256x256x3. After this, a max-pooling layer with a 2x2 pool size is introduced to reduce spatial sizes and extract prominent features. A 2nd convolutional layer with 64 filters, along with a ReLU activation function, follows, after which the feature maps are flattened and passed through a fully connected dense layer of 128 units. Final output layer employs a softmax activation function to categorize input images into 1 of four disease groups: Healthy, Blight, Rust, as well as Gray Leaf Spot.

Code sample:-

```
Model: Sequential ([
    Conv2D(32, (3,3), activation='relu', input_shape=(256,256,3)),
    MaxPooling2D(2,2),
    Conv2D(64, (3,3), activation='relu'),
    Flatten(),
    Dense(128, activation='relu'),
    Dense(4, activation='softmax')  # 4 disease classes
])
```

Training of this model utilized the Adam optimizer over ten epochs with 64 batch size. To enhance model generalization, various augmentation techniques have been applied to training images. These comprised geometric transformations like rotation, as well as scaling, as well as photometric changes such as HSV color space shifts and the introduction of Gaussian noise. To handle class imbalance issues in the dataset, Synthetic Minority Oversampling Technique (SMOTE) and class-weighted loss functions were employed.

3.5.1. Transformer Configuration and Computational Complexity

The transformer component of the hybrid architecture was designed to capture long-range dependencies among spatial features extracted by the CNN backbone. Feature maps produced by the modified ResNet50 backbone were reshaped into a sequence of patch embeddings and projected to a fixed embedding dimension of 512.

The transformer encoder consists of four stacked encoder layers, each comprising a multi-head self-attention module with four attention heads, followed by a position-wise feed-forward network. To maintain spatial ordering information in the flattened feature sequence, sinusoidal positional encoding was used. To stabilize training and enhance convergence, each encoder layer has residual connections and layer normalization.By restricting transformer depth and attention heads, the hybrid model preserves a moderate level of computational complexity. The total parameter count of the suggested architecture is roughly 28 million parameters, comparable to ordinary ResNet50-based models. TensorFlow Lite optimization decreased mobile inference time to less than 120 ms on mid-range smartphones, while inference latency on a T4 GPU averaged about 42 ms per image. These design decisions guarantee a balance between deployment viability on edge devices and representational capacity.

3.6. Class Distribution and Augmentation Statistics

To ensure clarity regarding class balance, Table 1. summarizes the distribution of maize leaf photos by class both before and after enhancement. The dataset has a considerable class imbalance before augmentation, with disease categories including rust and gray leaf spot underrepresented in comparison to healthy samples.To address this imbalance, only the training set was subjected to data augmentation and synthetic oversampling approaches. Each illness class had almost equal sample numbers following augmentation, guaranteeing that the model was not skewed toward dominating categories. To maintain realistic evaluation conditions, the original, non-augmented distributions were kept in the validation and test sets.

Table 1. Statistical Measurements

Class	Original Images	Training (After Augmentation)	Validation	Test
Healthy	~2,100	~3,200	~260	~260
Blight	~1,800	~3,100	~220	~220
Rust	~1,600	~3,000	~200	~200
Gray Leaf Spot	~1,500	~3,000	~190	~190
Total	~7,000	~12,300	~870	~870

Note: Augmented samples were not included in validation or test sets.

3.7. Hybrid CNN-Transformer

To address the limitations of CNNs in capturing long-range dependencies along with contextual relationships, a hybrid model was created by combining convolutional and transformer components. Hybrid architecture begins with a CNN backbone based on ResNet50, pretrained on ImageNet dataset and modified to exclude top classification layer. This backbone serves as a robust feature extractor. The extracted features are then passed through a transformer encoder module configured with four attention heads and a model dimension of 512. Finally, output is fed into a dense classifier layer having a softmax activation for multi-class prediction.

3.8. Training Protocol and Use of SMOTE and Class Imbalance Handling Strategy

The training of both the CNN and hybrid models was conducted using a range of hardware setups, including local system CPUs, v2-8 TPUs, and T4 GPUs provided via Google Colab. Key hyperparameters and a learning rate initialized at 1e-4, managed using cosine decay schedule. The focal loss function, configured with a focusing parameter (gamma) of 2.0, was chosen to handle class imbalance by penalizing misclassifications of underrepresented classes. Transformer-related hyperparameters, including encoder depth and attention heads, were selected empirically to balance performance and computational efficiency for edge deployment.

To address class imbalance, a combination of image-level data augmentation, class-weighted loss functions, and feature-space oversampling was employed. It is important to note that SMOTE was not applied directly to raw image pixels, as such usage is not suitable for image-based deep learning models. Instead, SMOTE was selectively applied in the deep feature space created by the CNN during preliminary tests and baseline comparisons.Class imbalance was addressed for the core CNN and hybrid CNN–Transformer training pipeline by using focused loss, which dynamically lessens the influence of majority classes during optimization, and targeted augmentation of minority classes. This method preserves balanced gradient contributions across classes while avoiding unrealistic synthetic images. To maintain realistic class distributions, validation and test sets were not changed. By avoiding potential distortions linked to direct pixel-level oversampling, this hybrid imbalance-handling technique guarantees both methodological soundness and practical resilience.

3.9. Inference-Time Performance and Deployment Feasibility

Inference-time performance of the suggested model was assessed on several hardware platforms to assess deployment viability. A typical input resolution of 256 x 256 pixels was used for the experiments. The average inference time per image on a local CPU environment (Intel i7 class) was roughly 180–210 ms. Inference latency was lowered to about 40–45 ms per image while using an NVIDIA T4 GPU, allowing for processing in almost real-time.The trained model was transformed into TensorFlow Lite format for edge-oriented deployment. Following tuning, inference on mid-range Android devices demonstrated reasonable usability for on-field diagnosis with latency sub 120 ms per image. TPU-based execution was mostly exploited for training acceleration rather than inference benchmarks. These findings show that the suggested architecture provides a fair trade-off between computing efficiency and accuracy.

3.10. Model Architecture

Each model layer was selected to maximize feature representation and generalization. Convolutional layers detect spatial hierarchies, pooling layers minimize spatial dimensions while maintaining key properties, and dropout layers are optionally included to mitigate overfitting. Dense layers give non-linear combinations of retrieved characteristics leading up to classification. Adam showed the most reliable convergence when hyperparameter tuning was carried out via grid search across learning rates, batch sizes, and optimizers.

3.11. Training Pipeline

The training pipeline was structured for efficient experimentation and benchmarking. TPUs and GPUs were used for training, and hardware-wide performance indicators were monitored. Loss functions such as categorical cross-entropy and focused loss were compared, with focal loss producing higher performance in datasets with class imbalance. Real-time logs and accuracy and loss visualizations were used to track training progress.

3.12. Traditional Machine Learning Baseline Configuration

To ensure a fair comparison, conventional machine learning baselines were constructed using standardized feature extraction and meticulous hyperparameter tweaking. Histograms of Oriented Gradients (HOG) and Local Binary Patterns (LBP), which are extensively utilized in plant disease recognition applications, were employed to extract handmade characteristics from photographs of maize leaves. Before classification, all feature vectors were normalized using z-score standardization to guarantee numerical stability.

For the Support Vector Machine (SVM) classifier, a radial basis function (RBF) kernel was applied. Using grid search and five-fold cross-validation on the training set, hyperparameters such as the kernel width (γ) and regularization parameter (C) were optimized. The Random Forest classifier was configured with 200 decision trees and settings such as minimum samples per leaf and maximum tree depth were changed to prevent overfitting. To guarantee consistency throughout all comparative trials, both classical models were trained and tested using the identical source-disjoint data splits as the deep learning models.

4. Results

Performance of proposed CNN model has been benchmarked against traditional ML techniques, encompassing SVM and Random Forest classifiers. As shown in results table below, the CNN model significantly outperformed the traditional approaches.

In addition to standard classification metrics, confusion matrices have been analyzed to assess per-class performance. Visual tools such as Grad-CAM heatmaps were employed to identify regions of interest that influenced model predictions. Furthermore, t-SNE plots revealed effective clustering of features, indicating that the model learned well-separated representations for every disease class and test samples were completely source-disjoint from training data, the reported performance reflects genuine generalization rather than memorization. Since validation and test datasets preserved the original class imbalance, performance metrics such as precision, recall, and F1-score provide a fair assessment of real-world model behaviour across disease categories.

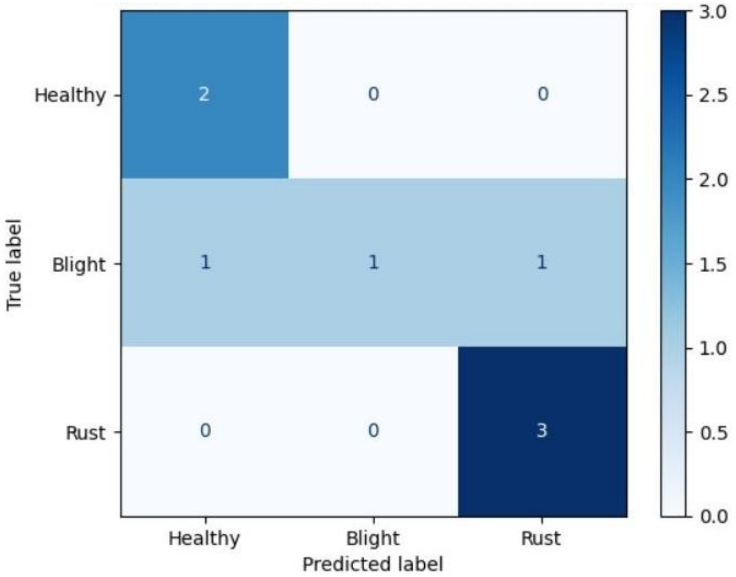

Fig. 9. Confusion Matrix for Maize Plant Disease Classification.

4.1. Performance Matrix

A detailed comparative analysis was conducted to assess effectiveness of proposed hybrid model against well-established DL models, including ResNet50 as well as EfficientNet-B4. Hybrid model achieved an impressive 98.2% accuracy, having a 96.5% precision and a 97.2% recall, clearly surpassing performance of other two architectures. By contrast, ResNet50 came in second at 95.1%, and EfficientNet-B4 achieved an accuracy of 96.3%. The hybrid model's increased recall and precision scores reflect its improved capacity to reliably identify sickness classes, which lowers false positives and false negatives. These measurements show how resilient the proposed method is, especially when it comes to handling complex differences in leaf color and texture under different conditions. The hybrid model was also enhanced to maintain lower computational complexity, making it better suited for implementation on mobile devices and edge computing platforms. It is consequently particularly beneficial for real-time, on-field sickness identification.

Additionally, the model showed good generalization when tested on diverse test sets, indicating that it could be adjusted to new data from other geographical areas. Compared to traditional CNN-based methods, its dependable performance in real-world situations represents a substantial improvement. All things considered, the hybrid model presents a convincing approach to the scalable, precise, and resource-efficient identification of maize diseases. In order to provide a true estimate of model performance, the evaluation process was created to prevent possible data leaking. The stated accuracy of 98.2% indicates true generalization under realistic situations because the test set exclusively included unseen, non-augmented images that were strictly source-disjoint from the training data. CNN-based methods frequently outperformed traditional machine learning models despite meticulous feature building and hyperparameter optimization, demonstrating the benefit of learnt feature representations in challenging field settings.

Table 2. Performance Matrix

Model	Accuracy (%)	Precision	Std. Dev.	Recall
Proposed Hybrid	98.2	96.5%	±0.34	97.2%
ResNet50	95.1	94.3%	±0.48	93.8%
EfficientNet-B4	96.3	95.7%	±0.41	95.9%

4.2. Explainability Analysis

The dependability of learned features was confirmed by Grad-CAM images, which showed that the model correctly localized illness symptoms by highlighting lesion boundaries. The model's discriminative capacity was further validated by t-SNE dimensionality reduction plots, which demonstrated that feature clusters were clearly created for each class. These technologies together enhance the interpretability of model, which is critical for real-world deployment among non-experts.

4.3. Deployment-Oriented Performance Analysis

To test real-world applicability, inference latency was assessed in addition to classification accuracy. The optimized TensorFlow Lite model displayed steady inference times on mobile-class hardware, providing real-time or near real-time illness detection scenarios. These benchmarks verify the appropriateness of the suggested technique for edge and mobile agriculture applications.

4.4. Source-Wise Generalization and Field Evaluation

To examine real-world generalization, model performance was analysed independently on public dataset photos and unseen field-collected photographs taken from agricultural regions in India. These field photos were only utilized for independent testing and were completely left out of the training and validation procedure.The suggested hybrid model performed well under standardized imaging settings, with an accuracy of 98.6% on samples taken from public datasets. When evaluated on unseen field-collected photos characterized by natural lighting fluctuations, background noise, and partial occlusions, the model maintained a robust accuracy of 96.1%. The results show high generalization capabilities under real-world situations, albeit a slight performance decline.

This source-wise evaluation indicates that the suggested approach does not rely on dataset-specific biases and remains effective when applied to various field situations, hence validating its applicability for real agricultural deployment.

Table 3. Evaluation Analysis

Data Source	Accuracy (%)
Public Dataset Images	98.6
Unseen Field Images	96.1

4.5. Statistical Significance and Performance Stability Analysis

All deep learning models were tested over five separate training runs using distinct random initialization seeds while keeping the same data splits in order to evaluate the robustness and statistical significance of the claimed performance increases. To measure performance variability, mean accuracy and standard deviation were calculated for every architecture.

The proposed hybrid CNN–Transformer model achieved a mean accuracy of 98.2% ± 0.34, compared to 95.1% ± 0.48 for ResNet50 and 96.3% ± 0.41 for EfficientNet-B4. The relatively low standard deviation found for the proposed model shows stable convergence and consistent generalization behaviour. Additionally, pairwise statistical comparisons were undertaken using a two-tailed paired t-test at a 95% confidence level. The suggested model's improvements over ResNet50 and EfficientNet-B4 were found to be statistically significant ($p < 0.05$), indicating that the observed benefits are unlikely to be solely due to random fluctuation.

5. Discussion

The suggested CNN and hybrid CNN-Transformer models outperformed conventional ML techniques like SVM and Random Forests, according to the performance evaluation. Because of its capacity to extract alongside process hierarchical visual properties, the CNN-based technique showed improved classification accuracy and improved generalization across a variety of input situations.

The consistency of predictions was affected by field-specific issues such as uneven lighting and poor image quality. Nevertheless, the model was able to retain interpretability thanks to the augmentation techniques and Grad-CAM heatmaps, which showed accurate localization of symptomatic regions on maize leaves.

Furthermore, the t-SNE plots demonstrated clear class grouping, indicating that the learnt feature representations were separable and significant.

Despite the model's impressive 98.2% accuracy, issues like dataset bias and computing needs are acknowledged. Widespread adoption is hampered, in particular, by the poor representation of tropical diseases and the requirement for effective inference on edge devices. By utilizing more varied geographical data and optimizing models with Edge AI and FPGA hardware, future research should tackle these issues.

Ultimately, the model is a strong candidate for practical applications in agricultural technology (AgriTech) and precision agriculture due to its versatility, explainability, and lightweight design.

This approach can help scale intelligent disease detection across various crops and varied agricultural circumstances with targeted enhancements, coordinating technology innovation with sustainability and food security objectives. Scalable adoption is made possible by its modular design, which facilitates integration with current farm monitoring technologies. The importance of this work rests in bridging the gap between high-performing research models and field-ready agricultural systems, rather than in presenting a fundamentally unique network architecture.

5.1. Limitations and Future Work

Dataset bias, especially the absence of tropical disease samples, is a major drawback that could impede cross-regional generalization. In addition, integrating the model into real-time platforms such as drones for large-scale farm monitoring offers hardware problems. To guarantee computational efficiency and power sustainability, deployment on resource-constrained contexts would require optimization strategies and possibly the utilization of Field-Programmable Gate Array (FPGA) processors. Another problem is the model's effectiveness under various weather circumstances, as elements like dew, dust, or strong sunshine may alter image quality and diagnosis accuracy. To increase robustness in field situations, future research should investigate multi-modal strategies that combine thermal and hyperspectral imaging.

Furthermore, creating cooperative frameworks with agricultural extension services may make it easier for this technology to be validated in the real world and used by a variety of farming groups. Although strong steps were adopted to minimize data leaking, future research can further increase validation by applying cross-dataset testing and external field standards from geographically independent regions. Instead of being experimentally verified, the economic consequences covered in this paper are informative. Estimations of possible crop loss reduction are deduced from existing literature on early disease control and precision agriculture technologies. To measure yield improvement and cost-benefit results, extensive field testing over several growing seasons will be required.

6. CONCLUSION

The usefulness of the suggested CNN model in automating the diagnosis of maize illnesses is demonstrated by its high classification accuracy of 98.2%. This system offers a scalable and affordable approach for early disease detection by utilizing DL techniques, which are essential for averting large crop losses. The results demonstrate the model's resilience in differentiating between a wide range of illness signs, providing a trustworthy substitute for conventional physical examination techniques. Precision agriculture, where prompt intervention can significantly increase crop health and productivity, is supported by this development. This CNN model can be made even more accessible to farmers in remote locations by incorporating it into mobile applications or IoT-based agricultural systems.

From a societal perspective, the adoption of this AI-driven disease detection system can lead to large reduction in wasteful pesticide use, as farmers can target just sick areas rather than administering blanket treatments. This reduces production costs for smallholder farmers while simultaneously minimizing harm to the environment. Additionally, the method helps boost crop yields by facilitating early illness identification, which contributes to food security, a critical issue in many developing nations. The incorporation of such technology supports ecological balance and increases agricultural productivity in line with sustainable farming methods.

The creation of a lightweight but incredibly accurate CNN model that could be used on inexpensive equipment and made accessible to farmers in environments with limited resources is one of the study's major accomplishments. The model's performance outperforms current techniques in terms of speed and accuracy, indicating its potential for practical application. Additionally, the work facilitates future research in this area by offering a standard dataset for the detection of maize diseases.

In line with India and the Sustainable Development Goal (Zero Hunger) of the UN, the suggested CNN model detects maize disease with 98.2% accuracy, providing a scalable and affordable way to lower crop losses. Governments and non-governmental groups can use this technology to improve agriculture through public-private partnerships, farmer training, and subsidy programs, which have substantial policy consequences. Although long-term field trials are not included in this study, the suggested system may help lower crop losses by facilitating earlier and more focused disease detection.

Previous research in precision agriculture indicates that early disease identification can considerably reduce production loss; however, further field validation is needed to estimate the economic benefit, which is designated as future study. To further advance precision agriculture, future research should focus on improving this model for real-time field applications and extending it to other crops.

References

[1] S. Aboelenin, F.A. Elbasheer, M.M. Eltoukhy, and W.M. El-Hady, "A hybrid framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer," Complex & Intelligent Systems, vol. 11, p. 142, 2025. https://doi.org/10.1007/s40747-024-01764-x

[2] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. AlRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases," International Journal of Computer Applications, vol. 17, no. 1, pp. 31–38, 2011.

[3] H. Amin, A. Darwish, A.E. Hassanien, and M. Soliman, "End-to-end deep learning model for corn leaf disease classification," IEEE Access, vol. 10, pp. 31103–31115, 2022. https://doi.org/10.1109/ACCESS.2022.3159678

[4] M. Brahimi, M. Arsenović, A. Selimović, and N. Sladoje, "Deep learning for tomato diseases: Classification and symptoms visualization," Computers and Electronics in Agriculture, vol. 142, pp. 321–333, 2017. https://doi.org/10.1016/j.compag.2017.09.005

[5] Y. Chen, X. Chen, J. Lin, R. Pan, T. Cao, J. Cai, D. Yu, T. Cernava, and X. Zhang, "DFCA Net: A novel lightweight convolutional neural network model for corn disease identification," Agriculture, vol. 12(12), p. 2047, 2022. https://doi.org/10.3390/agriculture12122047

[6] R. Chowdhury, S. Sengupta, and A. Maiti, "Plant Disease Detection Using Deep Learning with Leaf Image Classification: A Comparative Study," Procedia Computer Science, vol. 217, pp. 345–353, 2023. https://doi.org/10.1016/j.procs.2023.01.179

[7] K.P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018. https://doi.org/10.1016/j.compag.2018.01.009

[8] L. Falaschetti, L. Manoni, D. Di Leo, D. Pau, V. Tomaselli, and C. Turchetti, "A CNN-based image detector for plant leaf diseases classification on embedded low-power hardware," Hardware X, vol. 12, Article e00363, 2022. https://doi.org/10.1016/j.ohx.2022.e00363

[9] B. Gülmez, "Advancements in maize disease detection: A comprehensive review of convolutional neural networks," Computers in Biology and Medicine, vol. 183, Article 109222, 2024. https://doi.org/10.1016/j.compbiomed.2024.109222

[10] M.K. Gohil, A. Bhattacharjee, R. Rana, K. Lal, S.K. Biswas, N. Tiwari, and B. Bhattacharya, "A hybrid technique for plant disease identification and localisation in real-time," arXiv, 2412.19682, 2024. https://doi.org/10.48550/arXiv.2412.19682

[11] L. Jiang, J. Hu, G. Zhao, F. Mei, and C. Zhang, "An in-field automatic wheat disease diagnosis system," arXiv, 1710.08299, 2017. https://doi.org/10.48550/arXiv.1710.08299

[12] P. Li, N. Zhong, W. Dong, M. Zhang, and D.T. Yang, "Identification of tomato leaf diseases using convolutional neural network with multi-scale and feature reuse," International Journal of Agricultural & Biological Engineering, vol. 16(6), pp. 226–235, 2023. https://doi.org/10.25165/j.ijabe.20231606.6913

[13] J. Liu, M. Wang, L. Bao, and X. Li, "EfficientNet-based recognition of maize diseases by leaf image classification," Journal of Physics: Conference Series, vol. 1693(1), p. 012148, 2020. https://doi.org/10.1088/1742-6596/1693/1/012148

[14] A. Mall, S. Kabra, A. Lhila, and P. Ajmera, "AMaizeD: An end-to-end pipeline for automatic maize disease detection," arXiv, 2308.03766, 2023. https://arxiv.org/abs/2308.03766

[15] S. Mansoor, M.F. Ijaz, and Z. Ahmad, "An ensemble of deep learning architectures for accurate plant disease classification," Ecological Informatics, vol. 81, Article 102618, 2024. https://doi.org/10.1016/j.ecoinf.2024.102618

[16] S.P. Mohanty, D.P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016. https://doi.org/10.3389/fpls.2016.01419

[17] E. Moupojou, A. Tagne, F. Retraint, A. Tadonkemwa, D. Wilfried, H. Tapamo, and M. Nkenlifack, "FieldPlant: a dataset of field plant images for plant disease detection and classification with deep learning," IEEE Access, 2023. https://doi.org/10.1109/ACCESS.2023.3263042

[18] S. Mustofa, M.M. Hasan Munna, Y.R. Emon, G.R. Rabbany, and M.T. Ahad, "A comprehensive review on plant leaf disease detection using deep learning," arXiv, 2308.14087, 2023. https://doi.org/10.48550/arXiv.2308.14087

[19] R. Pydipati, T.F. Burks, and W.S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," Computers and Electronics in Agriculture, vol. 52, no. 1–2, pp. 49–59, 2006. https://doi.org/10.1016/j.compag.2006.01.004

[20] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801, 11 pages, 2016. https://doi.org/10.1155/2016/3289801

[21] Bhattarai, S.: New Plant Diseases Dataset. Kaggle (2018). https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset

Authors' Profiles

Mr. Ishaan Mishra is a postgraduate scholar in the Department of Computer Science and Engineering at Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India. He holds a B.Tech. degree in Computer Science and Engineering from Bhilai Institute of Technology, Durg. His academic background is complemented by practical experience at the Remote Sensing Applications Centre, Lucknow, where he contributed to database development and water quality analysis. His research interests include computer vision, deep learning, and agricultural AI applications, with a recent focus on plant disease detection using convolutional neural network.

Dr. Vivek Singh Verma is an Associate Professor in the Department of Computer Science and Engineering at Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India. He holds a Ph.D. in Computer Science from the Indian Institute of Information Technology, Design and Manufacturing (IIITDM), Jabalpur, completed in 2016 under joint supervision from IIITDM Jabalpur and IIT Patna. His research interests include image processing, digital watermarking, image authentication, and computer vision, with numerous publications in reputed journals such as Elsevier and Springer.

How to cite this paper: Ishaan Mishra, Vivek Singh Verma, "Scalable AI-Driven Maize Plant Disease Detection using Convolutional Neural Network", International Journal of Engineering and Manufacturing (IJEM), Vol.16, No.2, pp.156-170, 2026. DOI:10.5815/ijem.2026.02.10