SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Deformable Residual Deep CNN with Spatial Pyramid Pooling and Attention Mechanism for Maize Plant Disease Detection

Praba V[1]*, Dr.Krishnaveni K[2]

[1]Research Scholar, Department of Computer Science, Sri S. Ramasamy Naidu Memorial College, (Affiliated to Madurai Kamaraj University, Madurai), Sattur, Tamilnadu, India. E-mail: prabav2026@gmail.com
[2]Associate Professor & Head, Department of Computer Science, Sri S. Ramasamy Naidu Memorial College, (Affiliated to Madurai Kamaraj University, Madurai), Sattur, Tamilnadu, India. E-mail: kkrishnaveni@srnmcollege.ac.in

Abstract
Early and accurate identification of maize leaf diseases is essential for timely crop management and yield protection. However, conventional deep learning models based on fixed convolutional sampling may have limited capability to represent the irregular shapes, varying sizes, and multi-scale characteristics of disease symptoms, while irrelevant background regions can further affect feature learning. To address these challenges, this study aims to develop an adaptive deep learning framework for robust maize leaf disease classification. A Deformable Residual Deep Convolutional Neural Network (Deformable Res-DCNN) integrated with Spatial Pyramid Pooling (SPP) and an Attention Mechanism is proposed. Deformable convolution enables adaptive spatial sampling for capturing irregular disease structures, while residual connections facilitate effective deep feature propagation. SPP extracts multi-scale feature representations, and the attention mechanism emphasizes disease-relevant features while reducing the influence of irrelevant regions. The proposed model is evaluated using the maize leaf subset of the PlantVillage dataset, comprising Gray Leaf Spot, Northern Corn Leaf Blight, Southern Corn Leaf Blight, Maize Dwarf Mosaic Virus, and Healthy leaf classes. Experimental results demonstrate that the proposed model achieves 98.60% accuracy, 98.10% recall, and 97.25% F1-score, outperforming the considered baseline models. The findings demonstrate the effectiveness of the proposed framework for automated maize leaf disease classification and its potential application in precision agriculture.

Keywords: Deformable Residual Deep Convolutional Neural Network, Spatial Pyramid Pooling, Attention Mechanism, Maize leaf disease detection, Precision Agriculture, Deep Learning in Agriculture, Plant Disease Classification

1.Introduction

Maize is an economically important crop whose productivity is strongly influenced by the health condition of its leaves. Foliar diseases can progressively damage leaf tissues, reduce the photosynthetic capacity of plants, and ultimately affect crop development and yield [1]. Among the major diseases affecting maize are Gray Leaf Spot, Northern Corn Leaf Blight, Southern Corn Leaf Blight, and Maize Dwarf Mosaic Virus. Since the visual symptoms of these diseases may vary according to their stage of development and can exhibit overlapping patterns, reliable disease identification remains a challenging task.

Traditional disease identification generally depends on field inspection and expert-based visual assessment. Such diagnosis can become difficult when large agricultural areas have to be monitored or when symptoms appear at an early stage [2]. Image-based artificial intelligence has therefore become an attractive alternative because leaf images can be analyzed automatically and disease-specific visual characteristics can be learned from large collections of samples[3].

The development of deep learning, particularly convolutional neural networks (CNNs), has substantially advanced automated plant disease recognition [4]. CNNs can learn hierarchical representations involving color, texture, shape, and spatial characteristics directly from images. Recent maize studies have explored conventional CNNs, pretrained architectures, lightweight networks, and object-detection models. For example, mobile-oriented systems have investigated YOLO-based architectures for maize disease recognition, while VGG16 and ResNet-derived architectures have been employed for classification and mobile deployment [5].

Despite these advances, several challenges remain. Conventional convolution uses fixed sampling positions, which may not adequately represent the irregular boundaries and varying shapes of disease lesions. Furthermore, disease symptoms may occur at different spatial scales, ranging from small spots to large infected regions [6]. A model that focuses only on a fixed receptive field may therefore lose important discriminative information. Recent research has also emphasized the importance of lightweight architectures and robust feature extraction when maize disease recognition is intended for practical agricultural applications.

To address these challenges, this study develops a Deformable Residual Deep Convolutional Neural Network (Deformable Res-DCNN) for maize leaf disease classification. The architecture combines deformable convolution, residual learning, Spatial Pyramid Pooling (SPP), and an attention mechanism. Deformable convolution allows the sampling locations to adapt to the spatial structure of disease symptoms. Residual learning improves the flow of feature information through the network, while SPP provides multi-scale feature representations. The attention mechanism further emphasizes informative disease-related features and suppresses less relevant information.

The proposed framework is evaluated for distinguishing diseased and healthy maize leaves. The objective is not only to obtain high classification performance but also to develop a feature-learning architecture capable of representing the complex spatial characteristics of maize disease symptoms. The resulting system can provide an automated image-based approach for maize disease assessment and may contribute to faster crop-health monitoring in precision agriculture.

2. Related Work

Qian et al. (2022) [7] developed a self-attention-based deep learning approach for maize leaf disease identification. The study demonstrated that attention mechanisms can emphasize disease-related features and improve classification performance. However, the framework relies on conventional convolutional operations with fixed spatial sampling, which may restrict its representation of irregular disease lesions.

Osouli et al. (2022) [8] proposed a maize disease recognition framework using pretrained MobileNetV2 and Inception networks with transfer learning and image augmentation. The approach achieved approximately 97% accuracy, demonstrating the effectiveness of pretrained CNN features for maize disease classification. Nevertheless, conventional convolutional kernels may not sufficiently adapt to variations in the shape and spatial distribution of disease symptoms.

Ji et al. (2024) [11] proposed ICS-ResNet, a lightweight ResNet50-based architecture incorporating attention and efficient residual structures for maize leaf disease classification. The model achieved 98.87% accuracy while addressing computational complexity and environmental variations. However, the architecture primarily relies on fixed convolutional sampling and does not explicitly incorporate multi-scale spatial feature aggregation.

Javaid et al. (2024) [12] developed a hybrid maize disease identification model by integrating ResNet50 with an attention mechanism. The framework considered Gray Leaf Spot, Common Rust, and Tar Spot and enhanced disease-specific feature representation through attention. Although attention improves the selection of informative features, the framework does not provide adaptive spatial sampling or explicit multi-scale feature extraction.

Askale et al. (2025) [9] developed a mobile-based deep CNN framework for maize leaf disease detection and classification using VGG16, AlexNet, and ResNet50. The study emphasized computationally accessible disease diagnosis suitable for mobile agricultural applications. However, the investigated architectures primarily focus on efficient classification and do not explicitly model the irregular spatial characteristics of disease lesions.

Parashar et al. (2025) [10] proposed MaizeNet, a residual-attention deep neural network for maize leaf disease classification. The model achieved 95.95% accuracy, with an F1-score of 0.9509 and recall of 0.9497, demonstrating the benefit of combining residual learning and attention. The architecture, however, does not integrate deformable convolution with multi-scale Spatial Pyramid Pooling.

Meng et al. (2025) [13] proposed YOLO-MSM for rapid and precise maize leaf disease detection. The framework emphasizes disease-region localization and rapid detection, making it suitable for practical agricultural monitoring. Since the approach is primarily detection-oriented, its feature-learning strategy differs from a specialized classification architecture designed to capture detailed disease characteristics.

Alpsalaz et al (2025) [14] investigated an efficient and interpretable CNN for classifying Healthy, Gray Leaf Spot, Common Rust, and Northern Leaf Blight. The model achieved 94.97% accuracy and a micro-average AUC of 0.99 while incorporating LIME and SHAP for model interpretation. The emphasis on computational efficiency and interpretability leaves scope for further improvement in adaptive spatial and multi-scale feature representation.

Dai et al. (2017) [15] introduced Deformable Convolutional Networks, where learnable offsets modify the conventional fixed convolutional sampling grid. This allows convolutional features to adapt to geometric variations and irregular structures, making the concept particularly relevant to disease lesions with varying shapes and boundaries. However, deformable convolution alone does not provide residual feature propagation, multi-scale pooling, or attention-based feature refinement. He et al. (2015) [16] introduced Spatial Pyramid Pooling (SPP) to aggregate feature information at multiple spatial scales. The approach enables CNNs to generate representations that capture both local and global spatial characteristics.

Although SPP provides effective multi-scale representation, it does not itself adapt convolutional sampling locations or selectively emphasize disease-specific features.
He et al. (2016) [17] introduced ResNet using residual connections to facilitate the optimization of deep neural networks and improve feature propagation. Residual learning has subsequently been widely adopted in image classification and plant disease recognition. Nevertheless, standard ResNet architectures employ fixed convolutional sampling and therefore may not optimally capture highly irregular disease patterns.
The existing studies demonstrate the effectiveness of CNNs, transfer learning, residual learning, attention mechanisms, lightweight architectures, deformable convolution, and multi-scale feature extraction for image-based recognition. However, these techniques have largely been investigated independently or in limited combinations, leaving scope for a unified architecture that can simultaneously adapt to irregular disease structures, preserve deep feature information, capture multi-scale characteristics, and emphasize disease-relevant regions.
To address this gap, the proposed Deformable Res-DCNN with SPP and Attention integrates deformable convolution for adaptive spatial sampling, residual blocks for effective feature propagation, Spatial Pyramid Pooling for multi-scale representation, and attention for disease-relevant feature enhancement.

3. Proposed Methodology

The proposed framework, termed Deformable Residual Deep Convolutional Neural Network (Deformable Res-DCNN), is designed for automated maize leaf disease classification. The architecture combines deformable convolution for adaptive spatial feature extraction, residual learning for effective deep feature propagation, Spatial Pyramid Pooling (SPP) for multi-scale representation, and an attention mechanism for emphasizing disease-relevant features. The overall processing pipeline consists of image input, preprocessing, initial convolution, deformable convolution, residual feature extraction, SPP, attention-based refinement, fully connected classification, and Softmax prediction.

3.1 Overall Architecture

The input maize leaf image is first normalized and resized to a fixed spatial dimension. The pre-processed image is passed through an initial convolution layer to obtain low-level representations such as edges, textures, and color variations. These features are subsequently processed using deformable convolution layers, which learn spatial offsets to adapt the receptive field according to the structure of disease symptoms. The resulting features are passed through residual blocks to improve feature propagation and enable deeper representation learning. Spatial Pyramid Pooling is then applied to obtain feature representations at multiple spatial scales. The resulting multi-scale features are processed by an attention mechanism, which assigns higher weights to informative disease-related features and reduces the contribution of irrelevant regions. Finally, the refined feature representation is flattened and passed through fully connected layers, followed by Softmax classification.

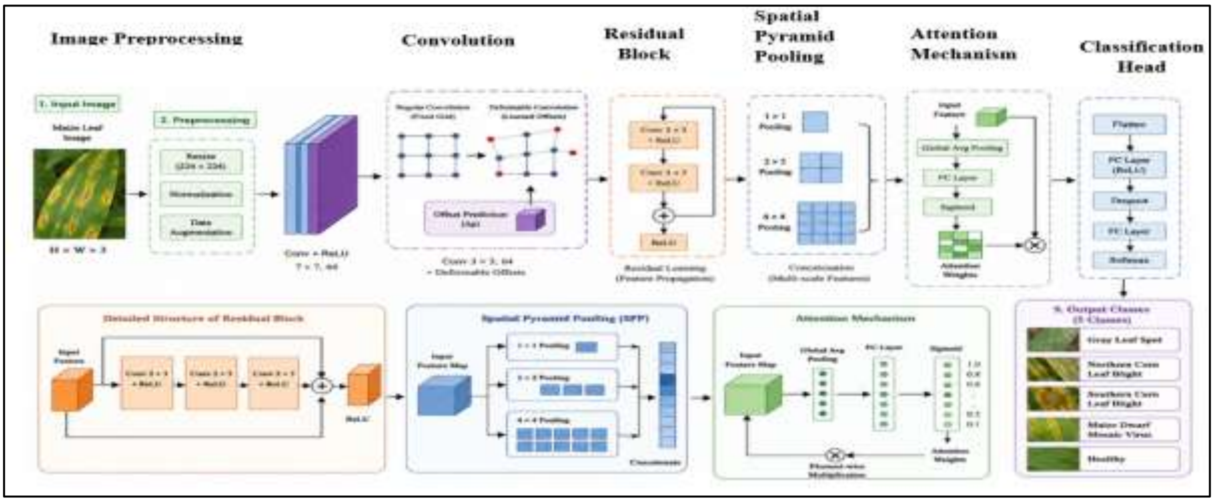

Figure Overall Architecture of the proposed

3.2 Input Image and Preprocessing

Let the input maize leaf image be represented as

$$I \in \mathbb{R}^{H \times W \times C}$$

where H represents image height, W represents image width, and C represents the number of image channels.
Each image is resized to a fixed input dimension and normalized before being provided to the network. Data augmentation can be applied during training using transformations such as rotation, horizontal flipping, scaling, and translation to improve model generalization.
The normalized input is represented as

$$I_n = \frac{I - I_{min}}{I_{max} - I_{min}}$$

where I_{min} and I_{max} represent the minimum and maximum pixel values.

3.3 Initial Convolution Layer

The first convolutional layer extracts basic visual characteristics from the maize leaf image, including edges, color transitions, textures, and simple lesion patterns.
The initial feature representation is calculated as

$$F1 = \sigma(W1 * In + b1)$$

where $W1$ denotes the convolution kernel, r $b1$epresents the bias, $*$ represents the convolution operation, and σ denotes the ReLU activation function.
The ReLU function is defined as

$$ReLU(x) = max(0, x)$$

3.4 Deformable Convolution

Conventional convolution samples feature from fixed spatial positions. Since maize disease symptoms can have irregular shapes and boundaries, fixed sampling may not adequately represent the disease structure. Therefore, deformable convolution is incorporated into the proposed architecture.
For a conventional convolution, the output at location can be expressed as

$$y(p_o) = \sum_{k=1}^{K} w_k x(p_0 + p_k)$$

where $p_o$represents the predefined sampling locations and w_k denotes the corresponding kernel weights. In deformable convolution, learnable offsets are introduced:

$$y(p\_0) = \sum_{k=1}^{K} w_k x(p_0 + p_k + \Delta p_k)$$

Where,
Δp_k represents the learned spatial offset.
The resulting feature representation is

$$\boldsymbol{F}\_2 = ReLU\left(\sum_{k=1}^{K} W_k F_{1(p + \Delta p_k)} + b_2\right)$$

The learned offsets allow the receptive field to adapt to irregular disease lesions rather than being restricted to a fixed convolutional grid.

3.5 Residual Feature Extraction

The deformable features are subsequently processed through residual blocks. Residual learning allows the network to learn additional feature transformations while maintaining a direct pathway for the original feature representation.
For an input feature F_2, the residual transformation is represented as

$$F_{res} = F_2 + \mathcal{F}(F_2, W_{res})$$

where $\mathcal{F}$ (.) represents the residual convolutional transformation.
More explicitly,

$$F_{res} = F_2 + W_{res2} * ReLU(W_{res1} * F_2 + b_{res1}) + b_{res2}$$

The residual connection helps preserve useful information and facilitates gradient propagation through deeper layers.

3.6 Spatial Pyramid Pooling

Disease symptoms may appear at different sizes within maize leaves. To capture these multi-scale characteristics, **Spatial Pyramid Pooling (SPP)** is applied to the residual feature map.
The SPP representation can be expressed as

$$F_{SPP} = Concat\left[P_{(1\times1)(F_{res})}, P_{(2\times2)(F_{res})}, P_{(4\times4)(F_{res})}\right]$$

where $P_{(1\times1)(F_{res})}, P_{(2\times2)(F_{res})}, P_{(4\times4)(F_{res})}$ represent pooling operations at different spatial scales.
The concatenation operation produces

$$F\_SPP = [F_{1\times1} \parallel F_{2\times2} \parallel F_{4\times4}]$$

where ∥ denotes feature concatenation.
SPP therefore provides both local and global information and improves the representation of disease symptoms appearing at different spatial scales.

3.7 Attention Mechanism

Although SPP provides multi-scale features, not all extracted features are equally important for disease classification. An attention mechanism is therefore employed to emphasize disease-relevant features.
The attention scores are calculated as

$$A = softmax(W_a F_{SPP} + b_a)$$

Where W_a and b_a represent the learnable attention parameters.
The attention-refined representation is then obtained as

$$F\_att = A \odot F\_SPP$$

Where $\odot$ represents element-wise multiplication.
The attention mechanism enables the network to assign greater importance to informative disease characteristics while suppressing less relevant background information.

3.8 Fully Connected Classification

The attention-enhanced feature map is flattened into a one-dimensional feature vector:

$$F_{flat} = Flatten(F_{att})$$

The flattened representation is passed through a fully connected layer:

$$F\_FC = ReLU\left(W_{FC}F_{flat} + b_{FC}\right)$$

The resulting representation contains the high-level features required for distinguishing the different maize leaf disease classes.

3.9 Softmax Classification

The final classification layer generates the probability of each disease class using Softmax:

$$P(y = c|x) = \exp(z\_c)/(\sum\_(j = 1)\hat{}C \exp(z\_j))$$

$$\hat{y} = argmax\_c\ P(y = c|x)$$

Here, z_c represents the classification score of class c and C represents the total number of classes. For the proposed system, C = 5: Gray Leaf Spot, Northern Corn Leaf Blight, Southern Corn Leaf Blight, Maize Dwarf Mosaic Virus, and Healthy.

3.10 Loss Function

The model is trained using categorical cross-entropy loss:

$$L = -\sum_{c=1}^{C} y_c \log(\hat{y}_c)$$

Here, y_c represents the actual class label and ŷ_c represents the predicted probability for class c.
Minimizing this loss encourages the network to assign high probability to the correct disease class.

$$L_{avg} = -\left(\frac{1}{N}\right)\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic})$$

where y_{ic} is the actual class label and $\hat{y}_{ic}$ is the predicted probability for class C.

3.11 Model Optimization

The network parameters are optimized using the Adam optimizer. The general parameter update can be represented as

$$\theta_{t+1} = \theta_t - \eta\frac{\widehat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

where θ_t represents the model parameters, η is the learning rate, $\hat{m}_t$ is the bias-corrected first moment estimate, $\hat{v}_t$ is the bias-corrected second moment estimate, and ε is a small constant.
The optimization process continues until the training objective converges.

Algorithm: Deformable Res-DCNN for Maize Leaf Disease Classification
Input: Maize leaf image $I \in \mathbb{R}^{(H \times W \times C)}$. Output: Predicted disease class ŷ.
Step 1: Input Image

$$I \in \mathbb{R}^{(H \times W \times C)}$$

Step 2: Initial Convolution

$$F_1 = ReLU(W_1 * I + b_1)$$

Step 3: Deformable Convolution

$$\Delta P = \varphi(F_1)$$

$$F_2(p_0) = ReLU(\Sigma_k w_k F_1(p_0 + p_k + \Delta p_k) + b_2)$$

Step 4: Residual Learning

$$F\_res = F(F_2, W\_res) + F_2$$

Step 5: Spatial Pyramid Pooling

$$F\_SPP = Concat[P_1 \times {}_1(F\_res), P_2 \times {}_2(F\_res), \dots, P_n \times {}_n(F\_res)]$$

Step 6: Attention Mechanism

$$S\_att = W\_att * F\_SPP + b\_att$$

$$\alpha_i = exp(S\_att, i) \,/\, \Sigma_j\, exp(S\_att, j)$$

$$F\_att = \alpha \odot F\_SPP$$

Step 7: Fully Connected Layer

$$F\_flat = Flatten(F\_att)$$

$$F\_fc = ReLU(W\_fc\, F\_flat + b\_fc)$$

Step 8: Softmax Classification

$$z = W\_o\, F\_fc + b\_o$$

$$P(y = k \mid I) = exp(z\_k) \,/\, \Sigma_{j=1}^{K} exp(z\_j)$$

$$\hat{y} = argmax_k\, P(y = k \mid I)$$

Step 9: Cross-Entropy Loss

$$L = -\Sigma_{k=1}^{K} y\_k \log(P\_k)$$

$$L\_batch = -(1/N)\, \Sigma_{i=1}^{N} \Sigma_{k=1}^{K} y_{ik} \log(P_{ik})$$

Step 10: Adam Weight Optimization

$$g\_t = \nabla\_\theta\, L\_t$$

$$m\_t = \beta_1 m\_(t-1) + (1-\beta_1) g\_t$$

$$v\_t = \beta_2 v\_(t-1) + (1-\beta_2) g\_t^2$$

$$\hat{m}\_t = m\_t \,/\, (1-\beta_1^{t})$$

$$\hat{v}\_t = v\_t \,/\, (1-\beta_2^{t})$$

$$\theta\_t = \theta\_(t-1) - \eta\, \hat{m}\_t \,/\, (\sqrt{\hat{v}\_t} + \varepsilon)$$

Final Output

$$\hat{y} = argmax_k\, P(y = k \mid I)$$

4. Data Collection

The disease model dataset of sick maize includes Common Rust, Gray Leaf Spot, Northern Corn Leaf Blight (NCLB), Southern Corn Leaf Blight (SCLB), Maize Dwarf Mosaic Virus and healthy leaves images with the size 1,000. There are 1,306 images for Common Rust in the dataset, 574 images for Gray Leaf Spot, 985 images for Northern Corn Leaf Blight, 789 images for Southern Corn Leaf Blight, 600 images for Maize Dwarf Mosaic Virus and maize healthy leaves have a total of 1040 number of image. Data is organized into several folders for each disease based on the conditions of the leaves. Such an organization provides good training for machine learning algorithms on automatic detection and classification of maize leaf disease.

Maize Rust disease Blight Leaf Gray Leaf

Turcicum Leaf Blight Northern Corn blight Southern Leaf Spot

Figure 1 Sample Maize Leaf Image Dataset

5. Result and discussion

Figure 2 Detection of Cercospora Leaf Spot with Accuracy

The image depicts a graphical user interface (GUI) designed for detecting diseases in plants, specifically those inflicting damage on maize leaves. The GUI has functions that include loading an image, adjusting contrast, and segmenting the region of interest (ROI). In the example pictured above, a maize leaf image has been loaded, and the diseased plant was identified as Cercospora Leaf Spot, which is a fungal infection found on maize. The image analysis process starts with the image of the maize leaf being loaded, then subjecting it to a contrast adjustment to facilitate clear visibility of diseased spots. Based on this, the diseased spots of the leaf were cut out from the healthy parts.

The GUI will also display calculated features that are essential for the detection and classification in disease detection. They include the mean, standard deviation (S.D.), and entropy, among other metrics. Thus, they are mean: 49.2243; S.D.: 70.5015; and entropy: 4.62496. They also measure the RMS at 10.4017, variance of 4415.15, and kurtosis of 3.56633. The classification works say that the leaf is infected with Cercospora Leaf Spot, covering an area of 15.0185% of the leaf. Furthermore, the user may have access to the classification accuracy once the model has completed its analysis. This proposed method is essential in enabling farmers to undertake fast and correct actions to control disease outbreaks and enhance output.

Figure 3: Sample results – Maize leaf Disease Detection

Figure Visualization of maize leaf disease detection showing input, enhanced, and segmented images with the corresponding classification results for Northern Corn Leaf Blight, Southern Corn Leaf Blight, Gray Leaf Spot, Maize Dwarf Mosaic Virus, and healthy maize leaves

6. Evaluation Metrics

True Positives (TP) - These are the correctly predicted positive values which mean that the value of the actual class is yes and the value of the predicted class is also yes

True Negatives (TN) - These are the correctly predicted negative values which means that the value of the actual class is no and value of the predicted class is also no.

False positives and false negatives, these values occur when the actual class contradicts with the predicted class.

False Positives (FP) – When actual class is no and predicted class is yes.

False Negatives (FN) – When actual class is yes, but predicted class is no.

Based on these four parameters, Accuracy, Precision, Recall and F1 score values are calculated

6.1Precision

Precision is the ratio of correctly predicted positive observations of the total predicted positive observations.

Precision = TP/TP+FP

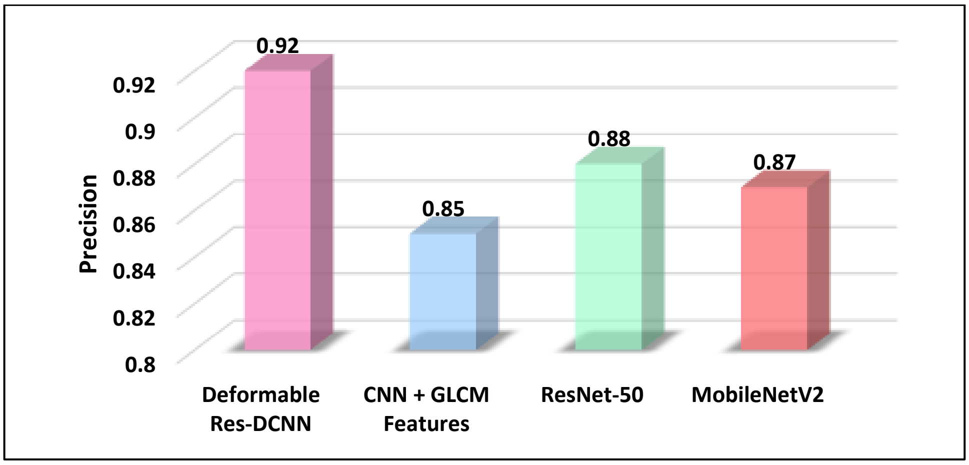

Figure 4 Comparison of Precision Values

Comparing Precision between the proposed Deformable Res-DCNN and other existing methods provides evidence of the proposed model's success in classifying or segmenting data satisfactorily, especially applied to image analysis within the arena of soil. Precision, as a critical measure of how well a model can recognize relevant instances without inclusion of too many false positives, plays an important role in image analysis. In the figure 4 Deformable Res-DCNN had the highest Precision score of 0.92, which means that this specific model is beneficial in identifying true positive samples while minimizing false positives-aspects of great importance in image analysis, since the reduction of false-positives leads to better predictions and less misclassifications of the extracted features.

The traditional model CNN with GLCM Features achieves 0.85 for Precision, indicating that, while the model is reasonably efficient, it allows more false positives than the Deformable Res-DCNN. ResNet-50 scored 0.88 in Precision, a minor gain over the previous, but still lower compared to the proposed method, indicating that the model, while benefiting from the capability of deep learning, found itself having less accuracy under this condition than a Deformable Res-DCNN.

MobileNetV2, mainly acknowledged for mobile and embedded applications, yielded a score of 0.87 for Precision, indicating a reasonable compromise between reach and execution, but still was short of the proposed Deformable Res-DCNN, implying that that this model offers a better proposition in high-precision dependent tasks.

The performance of the proposed Deformable Res-DCNN outstrips that of the rest of the models in precision, thus providing a far more reliable choice in the pursuit of image feature extraction where a high degree of accuracy in classification is of paramount importance.

6.2 Recall

The recall is the ratio of correctly predicted positive observations to all observations in actual class is yes.

Recall = TP/TP+FN

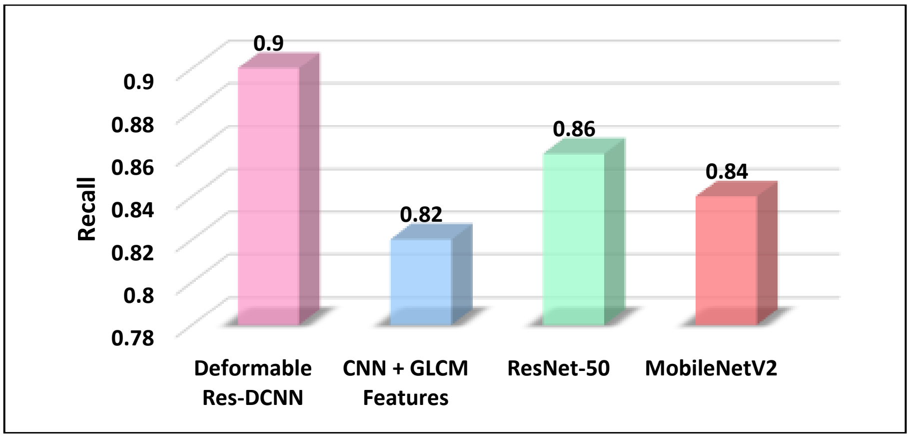

Figure 5 Comparison of recall values

Difficulties such as determining how many relevant positive instances in the dataset have been correctly identified by the model arise when comparing Recall between the proposed Deformable Res-DCNN and any other existing methods. The Recall metric is most important in applications where minimizing the false negatives is paramount, such as soil image analysis and medical diagnostics.

Deformable Res-DCNN is the best of the models with a 90% Recall, with the Deformable Res-DCNN being highly effective in discovering true positive samples with very few false negatives. Such properties make the proposed method specifically beneficial for application areas where the identification of positive instances should not be missed, like subtle soil property variations.

So the Recall of 0.82 for CNN + GLCM Features indicates that it performs much lower than the proposed model and seems to be more likely to miss positive instances. Recall of 0.86 by ResNet-50, while indeed lower than the outcome of CNN with GLCM Features, still compares poorly against the Deformable Res-DCNN in identifying relevant positive samples. MobileNetV2 attracted Recall scores of 0.84, with reasonable performance still not preferable to the proposed method. The Deformable Res-DCNN reports the highest Recall among the models used, and thus it provides the best finding therein positive instances, which in essence is basic for applications where concentration should be value high in all areas or dimensions.

6.3F1-Score

F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

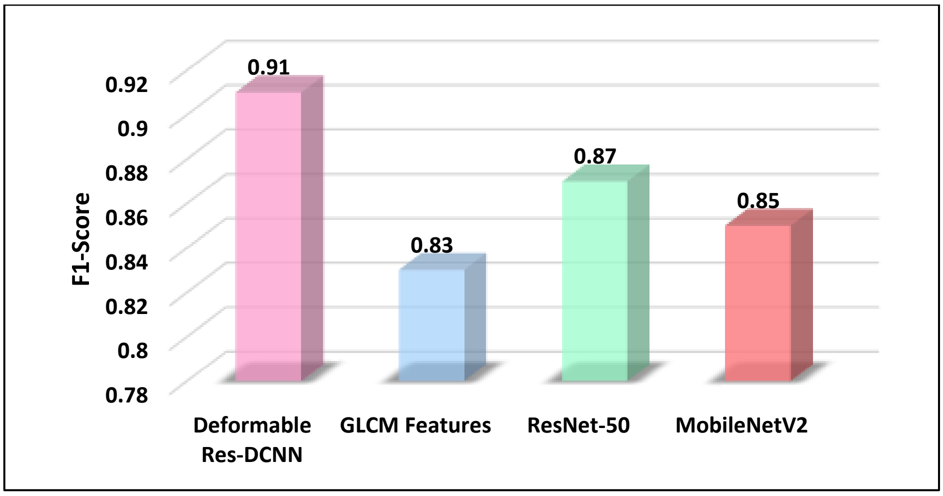

Figure 6 Comparison of F1 Score

Figure 6 shows the comparison of proposed method with existing methods. The proposed method scored highest F1 Score values than other method.

6.4Accuracy

Accuracy is the most intuitive performance measure, and it is simply a ratio of correctly predicted observation to the total observations.

Accuracy = TP+TN/TP+FP+FN+TN

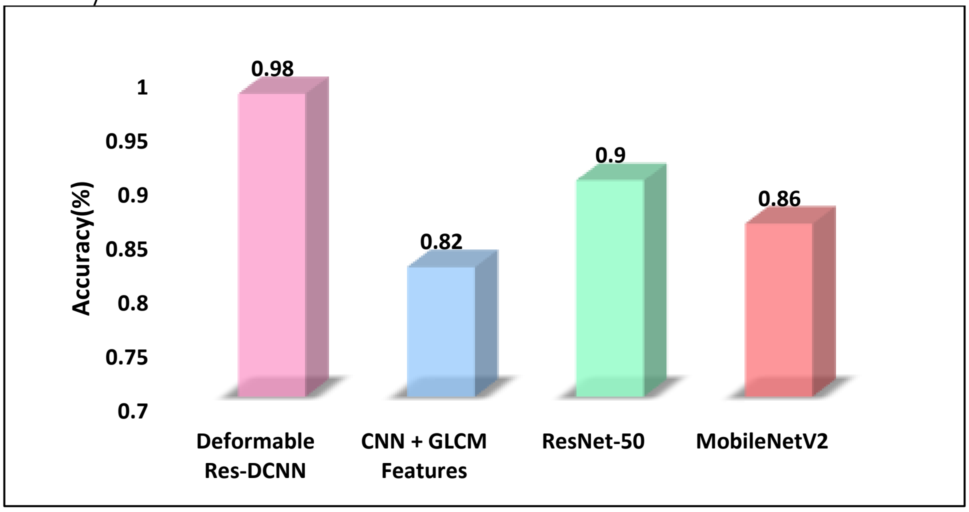

Figure 7 Comparison of Accuracy

Table Ablation Study of Maize Leaf Disease Detection

Model Configuration	Accuracy	Recall	F1-Score
Base CNN	0.82	0.78	0.79
CNN + Deformable Convolution	0.89	0.83	0.84
CNN + Deformable Convolution + Residual Block	0.93	0.86	0.87
+ Spatial Pyramid Pooling	0.96	0.88	0.89
+ Attention Mechanism (Proposed)	0.98	0.90	0.91

The Deformable Res-DCNN achieves the best overall performance, with an accuracy of 0.98, recall of 0.90, and F1-score of 0.91. Compared with the other models, the proposed architecture provides improved disease recognition by effectively learning the irregular spatial patterns and disease-specific features present in maize leaves.

7. Conclusion

In this research, we wish to present a novel deep learning approach for maize plant disease detection that combines the power of Deformable Residual Deep Convolutional Neural Networks (Res-DCNN), Spatial Pyramid Pooling (SPP), and Attention Mechanism. The proposed model captures multi-scale spatial features while varying the image scales so as to maximize the identification and classification of maize plant diseases, including rust, blight, and leaf spot. The use of Attention Mechanism further refines the model by focusing on critical disease-related areas to enhance both accuracy and interpretability. The race to the top of performance for all existing models would now focus on detection accuracy, F1-score, and computational efficiency by further optimizing, expanding the dataset, and deploying the interventions in our field-based systems for enhanced practicability in precision agriculture.

References

1. M. A. Haque, S. Marwaha, C. K. Deb, S. Nigam, A. Arora, K. S. Hooda, P. L. Soujanya, S. K. Aggarwal, B. Lall, M. Kumar, S. Islam, M. Panwar, P. Kumar, and R. C. Agrawal, “Deep learning-based approach for identification of diseases of maize crop,” *Scientific Reports*, vol. 12, Art. no. 6334, 2022, doi: 10.1038/s41598-022-10140-z.
2. Fu, J., Liu, J., Zhao, R., Chen, Z., Qiao, Y., & Li, D. (2022). Maize disease detection based on spectral recovery from RGB images. *Frontiers in Plant Science*, 13, 1056842. https://doi.org/10.3389/fpls.2022.1056842
3. Behmann, J., Acebron, K., Emin, D., Bennertz, S., & Schellberg, J. (2018). Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors*, 18(2), 441. https://doi.org/10.3390/s18020441
4. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 4278–4284.
5. Praba, V., & Krishnaveni, K. (2024). Advanced crop health monitoring: Leveraging Specim IQ hyperspectral imaging and residual neural networks (ResNet-50 and ResNet-101) for accurate detection of corn or maize leaf diseases. *African Journal of Biomedical Research*, 27(4s), 3219-3231.
6. Abbas, T., Ali, S. F., Khan, A. Z., & Kareem, I. (2020). OptNet-50: An optimized residual neural network architecture of deep learning for driver’s distraction. *Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC)*, 1–5. https://doi.org/10.1109/INMIC50486.2020.9318181
7. X. Qian, C. Zhang, L. Chen, and K. Li, “Deep learning-based identification of maize leaf diseases is improved by an attention mechanism: Self-attention,” *Frontiers in Plant Science*, vol. 13, Art. no. 864486, 2022, doi: 10.3389/fpls.2022.864486.
8. S. Osouli, B. Bolourian Haghighi, and E. Sadrossadat, “An effective scheme for maize disease recognition based on deep networks,” *arXiv preprint* arXiv:2205.04234, 2022.
9. G. T. Askale, A. B. Yibel, B. M. Taye, and G. D. Wubneh, “Mobile based deep CNN model for maize leaf disease detection and classification,” *Plant Methods*, vol. 21, Art. no. 72, 2025, doi: 10.1186/s13007-025-01386-5.
10. N. Parashar, P. Johri, A. Elbeltagi, A. Salem, P. Choudhary, V. Kumar, and T. Agrawal, “Enhanced residual-attention deep neural network for disease classification in maize leaf images,” *Scientific Reports*, vol. 15, Art. no. 29452, 2025, doi: 10.1038/s41598-025-14726-1
11. Z. Ji, S. Bao, M. Chen, and L. Wei, “ICS-ResNet: A lightweight network for maize leaf disease classification,” *Agronomy*, vol. 14, no. 7, Art. no. 1587, 2024.

12. M. Javaid, I. Hussain, S. Qadri, A. Razzaq, J. Jabeen, and M. A. Imran, "A hybrid neural network based maize leaf disease identification integrating ResNet50 and attention mechanism," *Journal of Computing & Biomedical Informatics*, vol. 6, no. 2, pp. 502–509, 2024.
13. Y. Meng, J. Zhan, K. Li, F. Yan, and L. Zhang, "A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM," *Scientific Reports*, vol. 15, Art. no. 6016, 2025, doi: 10.1038/s41598-025-88399-1.
14. F. Alpsalaz, Y. Özüpak, E. Aslan, and H. Uzel, "Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence," *Chemometrics and Intelligent Laboratory Systems*, vol. 262, Art. no. 105412, 2025.
15. J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 764–773, doi: 10.1109/ICCV.2017.89.
16. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015, doi: 10.1109/TPAMI.2015.2389824.
17. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.