Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Efficient Ensemble of Pre-Trained CNN Models for Improving Classification of Maize Diseases

Pankaj Raj Dawadi[1], Sumanta Silwal[2]

[1]Department of Computer Science and Engineering, School of Engineering, Kathmandu University, Dhulikhel, Nepal.

[2]Sunway College, Kathmandu, Nepal.

Email: [1]pdawadi@ku.edu.np, [2]sumantasilwal10@gmail.com

Abstract

Maize is one of the world's leading staples crops, whose yields are compromised by foliar diseases such as rust, blight, and gray leaf spot. Speedy and correct diagnosis must be employed to reduce loss and maintain food security. Deep learning or Convolutional Neural Networks (CNNs) has been promising for auto-classifying diseases, but one-model systems are not robust and generalizable. Moreover, previous research did not address issues of dataset imbalance leaf orientation variation in the out-worldly setting, and field deployability for precision farming. In this research, a feature ensemble-fusion classifier model for maize disease identification is presented. Pre-processing and augmentation were performed on a 4,188-leaf image dataset divided into four classes (Common Rust, Gray Leaf Spot, Blight, and Healthy). Six pre-trained CNNs (EfficientNetB0, DenseNet201, ResNet50V2, NasNetMobile, MobileNetV2, and VGG16) were tested in frozen and partially fine-tuned states. EfficientNetB0, DenseNet201, and ResNet50V2 were the top three models averaged using an ensemble approach. The proposed system achieved a total average of 98% accuracy across all diseases, with 97% precision, 96% recall, and 96.5% F1-score, and at least 92% precision, 96% recall, and 96% F1-score on Gray Leaf Spot disease thus performing better than single CNNs. With improved class imbalance handling and environmental robustness, the scheme offers an efficient and adaptive solution to real-world maize disease diagnosis and precision agriculture.

DOI: https://doi.org/10.36548/jaicn.2025.003
Received: 21.07.2025, received in revised form: 22.08.2025, accepted: 09.09.2025, published: 23.09.2025

Keywords: Maize Disease, Convolutional Neural Networks (CNNs), Transfer Learning, Ensemble Model, Disease Classification.

1. Introduction

Plant disease is a general worldwide threat to food security, inducing substantial losses in yield and financial resources. For example, even partial control over soybean rust could have saved farmers millions of dollars [1]. Maize, the commodity used for food, feed, and biofuel, is highly vulnerable to foliar pathogens such as Common Rust, Gray Leaf Spot, and Corn Leaf Blight, all of which are reported to cause severe effects on crop yields [2], [3]. Among these, the most common in the tropics are Corn Leaf Blight (Exserohilum turcicum) and Common Rust, the latter of which can lead to up to 50% yield loss if infection occurs early in growth [4].

The classical approaches to disease diagnosis based on visual examinations are time-consuming, subjective, and non-reproducible, resulting in faulty diagnoses and inappropriate use of chemical treatments. All these limitations imply the need for low-cost and reproducible automatic diagnostic tools for precision agriculture. Deep learning, and specifically Convolutional Neural Networks (CNNs), present a solution. CNNs have the ability to learn high-level visual features and are consequently very well adapted to image-based classification problems like plant disease detection [5], [6].

Figure 1. Corn Leaf Diseases: Left: Blight, Middle: Common Rust and Right: Gray Leaf Spot [7]

While all the above work involves single CNN models that are susceptible to limited generalizability, light and leaf orientation dependence, and are also challenged in dealing with dataset imbalance.

This study examines the application of CNNs for maize disease diagnosis using the comparison of six pre-trained models, i.e., VGG16, DenseNet201, EfficientNetB0, MobileNetV2, NASNetMobile, and ResNet50V2. Three methods are employed to identify the optimal diagnostic approach: (i) training individual classification layers with frozen convolutional bases, (ii) fine-tuning a small number of layers, and (iii) constructing an ensemble of top-performing models to increase robustness and precision. The ensemble model overcomes the research gap found in earlier work done with a single CNN only, and it offers accuracy and reliability. In addition, the design is light and scalable, and it can provide timely and correct disease diagnosis for improved crop protection and yield enhancement.

This paper is articulated as follows: Chapter 1 comprises the introduction and background. Chapter 2 is the review of relevant works, including the thematic area, supporting technologies, and a study of different CNN models that are widely used for maize disease classification. It further consists of a literature review of pre-trained CNN models along with a gap analysis. Chapter 3 explains the suggested ensemble model structure and dataset pre-processing for the pre-trained CNN models. Chapter 4 discusses the experimental outcomes and conclusions. Finally, Chapter 5 provides the conclusion, declares the limitations, and suggests future work directions.

2. Related Work

Grey Leaf Spot (GL), caused by the Cercospora zeae-maydis fungal disease, is one of the most destructive limitations to maize production globally. It is an epidemic disease in much of the eastern United States and has been spreading to broader areas over the last few years. Lower leaves typically show primary symptoms first [8]. Common Rust (CR), caused by Puccinia sorghi, thrives in cold (16–23°C), wet conditions and is photosynthetically inhibitory by infecting both upper and lower leaf surfaces [9]. Northern Corn Leaf Blight (NCLB), caused by Exserohilum turcicum, appears as long, grayish blight and is well-developed under mild and humid climatic conditions [10]. These types of studies are mostly deficient in maize specialization, balanced data, and field or growth records, making them less informative.

While there are laboratory methods of detection for leaf diseases available, they cannot be applied on a daily basis because of cost and time factors. Deep learning algorithms have been utilized to identify plant diseases optimally. OpenCV-based CNNs were employed by [11] with 91% to 98% accuracy for leaf disease classification. Additionally, [12] was trained on a dataset with more than 54,000 leaf images for 14 crops and 26 diseases, yielding 99.35% classification accuracy. These works assure the usability of CNNs, although most of the analyses are crop generic and not maize-specific.

Other methods have also been explored. [13] utilized artificial neural networks (ANNs) with spectral radiometer data to diagnose several stages of fungal infection in oil palms. [14] proposed a web-based model for pomegranate disease diagnosis, achieving an accuracy of 82%. [15] compared ANN and SVM classifiers using color and texture features, with SVM outperforming ANN with an accuracy of 92.17% compared to 87.4%. These methods are either computationally bound at inference time, crop-specific, or challenging to scale for real-world field applications.

Despite these potential improvements, there are new areas of research that need to be addressed. Most work relies on a single CNN model, lacking ensemble methods that combine results from multiple models to boost robustness and accuracy. Large datasets are often composed of multi-crop and imbalanced data, failing to give sufficient weight to maize leaves. Prior software has been experimental in nature, without consideration for actual deployment in the field, ease of use on edge devices, and operation under variations in lighting, background, and leaf orientation.

Therefore, this work addresses these gaps by proposing an ensemble of pre-trained CNN models for the exclusive diagnosis of maize diseases, aiming to identify three enhanced leaf diseases (NCLB, CR, GL) and separate them from a well-balanced, enhanced dataset with high robustness and realistic application.

3. Proposed Work

The entire experiment procedure discussed in this paper comprises three main steps: dataset preparation, ensemble model architecture of CNN, and performance comparison of the suggested ensemble approach.

3.1 Dataset Preparation

The database had 4,188 images of four classes: Blight (1,146), Common Rust (1,306), Gray Leaf Spot (574), and Healthy (1,162) and which were downloaded from [16]. Images were taken in the field in New Zealand to introduce natural variation in illumination, background, and leaf orientation. Preprocessing and augmenting the data included Gaussian blurring, resizing to 224×224 pixels, and CLAHE (contrast-limited adaptive histogram equalization).

Rotation, flipping, shearing, and brightness adjustment were utilized for data augmentation and class balancing for generalization. The data augmentation processes added variation to the training datasets. The dataset was also divided into training (3,196 images, 80%), validation (526 images, 10%), and test (572 images, 10%) sets based on stratified sampling to ensure class distribution.

3.2 Model Architecture, Training, and Testing

Convolutional Neural Networks (CNNs) were utilized because they can automatically learn from image features for image classification. To take advantage of existing knowledge and minimize training time, six pre-trained CNN models, namely VGG16, DenseNet201, EfficientNetB0, MobileNetV2, NASNetMobile, and ResNet50V2, were utilized because they have an extremely well-balanced depth, parameter efficiency, and high performance when leveraging the transfer learning.

Three training scenarios were explored:

- Frozen convolutional layers: The base model's convolutional layers were frozen, and a custom classification head with dropout layers was trained.
- Partial fine-tuning: The top 20% of convolutional layers were fine-tuned while 80% remained frozen, enabling task-specific learning.
- Ensemble of top-performing models: The top three models based on validation F1-Score were combined to improve robustness and balance performance trade-offs.

All models were trained with the Adam optimizer, a learning rate 0.0001, a batch size of 32, and up to 50 epochs with early stopping if the validation loss did not decrease over a specified number of epochs. They are referred to as Model A, Model B, and Model C, respectively, and the description of the models is provided in Table 1.

Table 1. Pretrained Models with Proposed Model

Model	Architecture
Model	VGG16, DenseNet201, EfficientNetB0, MobileNetV2, NasnetMobile,
Model	80% of the layers of the base mode + 20% available training
Model	Ensemble model

3.3 Proposed Ensemble Model

Based on Table 1, the six pre-trained models are trained in three environments in this work. Of the three top models ranked by the F1-Score metric, they were selected to be included in an ensemble. Among the various ensemble techniques, one was a concatenated feature-level ensemble. It is an ensemble technique that concatenates feature representations achieved by other models in a manner that allows the architecture to learn richer and more diverse features.

The ensemble approach takes advantage of the power of an ensemble of CNNs to generate improved prediction performance, efficiency, and scalability compared to the limitations of each individual CNN model. The top three pre-trained models (M_1, M_2, M_3) are mapped onto a feature vector for each input image in parallel. The feature vectors are summed as a unified representation and passed through two fully connected layers (512 → 256 units, ReLU activation) prior to the SoftMax classification layer.

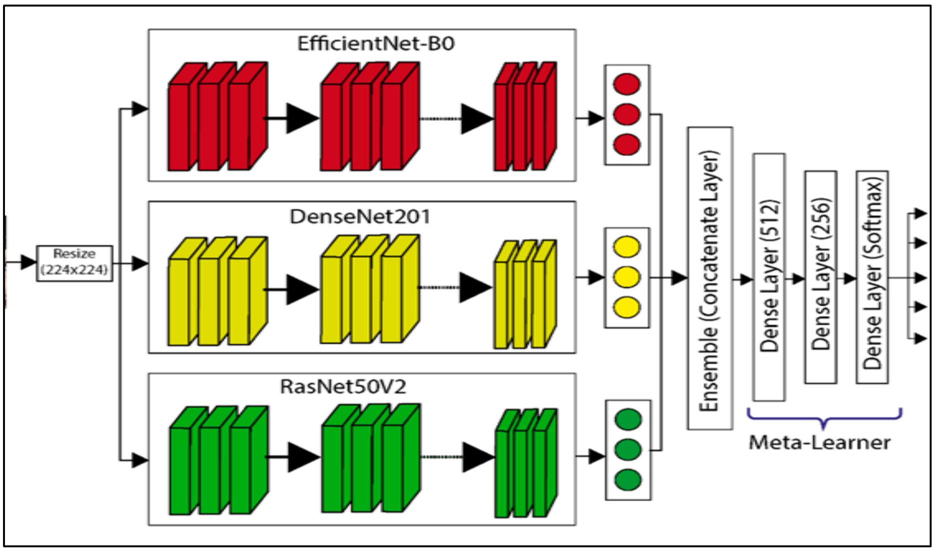

Figure 2. Proposed Ensembled Method

The ensemble approach reduces bias and variance effects, enhances class imbalance performance, and generalizes better with variations in illumination, background, and orientation of leaves. Model parameters are optimized using categorical cross-entropy loss and the Adam optimizer. Validation loss is tracked every epoch, and early stopping facilitates convergence in optimal time.

Figure 2. Proposed CNN-Based Ensemble Architecture for Maize Disease Classification. The three pre-trained CNNs M_1, M_2, and M_3 that are employed for maize leaf image feature extraction from the parallel inputs are dictated by the architecture. The feature vectors from the three models are concatenated and fed into fully connected layers (512 → 256 units) before being passed to the SoftMax classification layer. Annotated improved and frozen layers are the implementation of transfer learning, while the ensemble strategy enhances robustness and generalization against variations in lighting conditions, background, and leaf orientation.

The processing pipeline of the proposed ensemble approach is presented as pseudocode as follows:

- Start
- Load Pre-trained Models (M_1, M_2, M_3) → three boxes feeding into next stage
- Training Loop (Epoch t = 1…T)
- Batch Loop (i = 1…N_t)
- Input image → M_1, M_2, M_3 in parallel → F_1, F_2, F_3
- Concatenate → Dense(512, ReLU) → Dense(256, ReLU)
- Loss = CategoricalCrossEntropy
- Update Θ (Adam)
- Validation
- M_1(X_val), M_2(X_val), M_3(X_val) → Concatenate → Dense(num_classes) → Softmax
- Compute L_val
- Early Stopping Check (p steps without improvement?)
- Stop & Return Ensemble Model

4. Results and Discussion

4.1 Dataset and Experimental Setup

The maize disease image dataset comprised 4,188 images grouped into four classes: Blight (1,146), Common Rust (1,306), Gray Leaf Spot (574), and Healthy (1,162). The dataset was partitioned into training (3,196, 80%), validation (526, 10%), and testing (572, 10%) sets using stratified sampling to preserve class distributions.

Two training sets were compared for each CNN model:

Scenario 1: All the pre-trained convolutional layers were frozen; three additional fully connected layers with a SoftMax classifier were trained for 30 epochs at a learning rate of 0.0001.

Scenario 2: The highest 20% of convolutional layers were not frozen so that fine-tuning could take place with the pre-trained wide representations intact; training took place for 30 epochs as before.

This design ensured an equal comparison between models while filling gaps in existing research that, at times did not include a systematic evaluation of transfer learning methods.

4.2 Performance of Individual CNN Models

Some of the experiments were performed, i.e., six pre-trained CNN models (VGG16, DenseNet201, EfficientNetB0, MobileNetV2, NASNetMobile, ResNet50V2) under both training modes, their selectively frozen versions, and the developed ensemble model. Precision, recall, F1-score, and accuracy were used to evaluate the four classes of maize leaves. Table 2 presents the performance of individual CNN models. ResNet50V2, DenseNet201, and EfficientNetB0 attained a steady accuracy of over 96%, reflecting excellent performance in maize-specific disease characteristic detection.

Table 2. Experimental Results for Scenario 1 and Scenario 2 on the Image Dataset

CNN Models	Scenario	Precision	Recall	F1_Score	Accuracy (%)
VGG16	Sc. 1	0.91	0.92	0.91	93
	Sc. 2	0.07	0.25	0.11	28

DenseNet201	Sc. 1	0.94	0.93	0.94	95
	Sc. 2	0.96	0.96	0.96	97
EfficientNetB0	Sc. 1	0.94	0.94	0.94	95
	Sc. 2	0.97	0.98	0.98	98
NasnetMobileV2	Sc. 1	0.89	0.91	0.90	91
	Sc. 2	0.85	0.86	0.82	84
MobileNetV2	Sc. 1	0.90	0.91	0.90	92
	Sc. 2	0.92	0.93	0.92	93
ResNet50V2	Sc. 1	0.94	0.95	0.95	96
	Sc. 2	0.97	0.97	0.97	97

4.3 Ensemble Model Performance

The three top-performing models, DenseNet201, EfficientNetB0, and ResNet50V2, were selected based on validation F1-scores and combined into a bagging-based feature-level ensemble. The ensemble was trained for over 30 epochs, and hyperparameter tuning was applied to the four maize disease classes individually. Table 3 shows the performance metrics of the three top models and the developed ensemble, taking into consideration precision, recall, F1-score, and accuracy for all classes individually.

Table 3. Performance Measurement of Top Three Model and Proposed Method

Classes	Model	Precision	Recall	F1_Score	Accuracy (%)
Blight	EfficientNetB0	0.97	0.97	0.97	98
	DenseNet201	0.93	0.95	0.94	97
	ResNet50V2	0.94	0.97	0.96	97
	Proposed Model	0.97	0.96	0.96	98
Common_Rust	EfficientNetB0	0.99	0.99	0.99	98
	DenseNet201	0.99	0.99	0.99	97

	ResNet50V2	0.99	0.99	0.99	97
	Proposed Model	0.99	0.99	0.99	98
Gray_Leaf_Spot	EfficientNetB0	0.93	0.95	0.94	98
	DenseNet201	0.92	0.88	0.90	97
	ResNet50V2	0.94	0.91	0.93	97
	Proposed Model	0.92	0.96	0.94	98
Healthy	EfficientNetB0	1.00	0.99	0.99	98
	DenseNet201	1.00	0.99	1.00	97
	ResNet50V2	1.00	0.99	0.99	97
	Proposed Model	1.00	0.99	0.99	98

The ensemble performed better than single models all the time, particularly on Gray Leaf Spot, where visual similarity consistently led to misclassification by single models. This confirms that the feature-level ensemble can be successful in mitigating individual architectures' weaknesses and improving generalization to real-world variation.

4.4 Validation of Results

In an effort to further validate the ensemble, three-fold cross-validation was conducted. The data were partitioned into three equal folds and one was utilized once as validation while the other two were utilized for training. Table 4 presents mean F1-scores for all the folds.

Table 4. Three-Fold Cross Validation (Mean F1-Score)

CNN Models	**F1-Score**		
	Fold-1	**Fold-2**	**Fold-3**
EfficientNetB0	0.9732	0.9698	0.9788

	Mean F1-Score = 0.9739		
DenseNet201	0.9765	0.9699	0.9667
	Mean F1-Score = 0.9710		
ResNet50V2	0.9767	0.9689	0.9621
	Mean F1-Score = 0.9692		
Proposed method	0.9833	0.9866	0.9765
	Mean F1-Score = 0.9822		

The group achieved the highest and most consistent F1-scores for each of the folds because it established stability and possessed the ability to generalize exceptionally well.

The accuracy of the test ensures that the ensemble model constructed does in fact enhance maize disease classification more effectively than standalone pre-trained CNN models. EfficientNetB0, DenseNet201, and ResNet50V2 were the three best CNNs among the six evaluated, achieving the three best overall accuracies and equaling the best feature extraction performance in differentiating visually similar disease patterns. By feature level integration of these models, the system sidesteps the one-model limitation, recovers from mistakes, and works equally well on all four disease classes, even sophisticated ones such as Gray Leaf Spot. Three-fold cross-validation also identified the stability of the ensemble with the best mean F1-score (0.9822) and thus its generalizability. It surpasses single CNNs and previous multi-crop transfer learning, and is also invariant to in-the-wild lighting, leaf orientation, and background variation.

5. Conclusion

By addressing the drawbacks of conventional single-CNN model solutions, this article proposes a potent and effective CNN-based ensemble approach for maize disease classification. Based on the comparison of six pre-trained CNNs in frozen and partially fine-tuned conditions, the top three models—EfficientNetB0, DenseNet201, and ResNet50V2—were selected and then ensembled at the feature level. With an overall accuracy of 98%, the trained ensemble outperformed individual models and the state-of-the-art in recall and F1-

measure for each disease class. Three-fold cross-validation showed model compliance and generalization ability, even though ensemble learning and data augmentation ensured insensitivity to changes in background, illumination, and leaf orientation. Nevertheless, this method produces computational efficiency that is compatible with real-world field or mobile deployment applications. A reliable and scalable method for automated maize disease detection is presented in the paper, which could aid in timely intervention, lower crop loss, and increase agricultural productivity.

References

[1] Sankaran, Sindhuja, Ashish Mishra, Reza Ehsani, and Cristina Davis. "A review of advanced techniques for detecting plant diseases." Computers and electronics in agriculture 72, no. 1 (2010): 1-13.

[2] https://www.studocu.com/row/document/university-of-the-gambia/introduction-to-management/food-security-update-lxxii-october-27-2022/44177935

[3] Food and A. O. of the United Nations, “2021 ANNUAL REPORT, Plant Production and Protection,” 2021. http://old.belal.by/elib/fao/1436.pdf

[4] T. A. Jackson-Ziems, “Northern Corn Leaf Blight,” 2016, University of Nebraska–Lincoln Extension.

[5] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Communications of the ACM 60, no. 6 (2017): 84-90.

[6] LeCun, Yann. "Deep learning & convolutional networks." In 2015 IEEE hot chips 27 symposium (HCS), IEEE Computer Society, (2015): 1-95.

[7] Singh, Davinder, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. "PlantDoc: A dataset for visual plant disease detection." In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, (2020): 249-253.

[8] Ward, Julian MJ, Erik L. Stromberg, David C. Nowell, and Forrest W. Nutter Jr. "Gray leaf spot: a disease of global importance in maize production." Plant disease 83, no. 10 (1999): 884-895.

[9] Dillard, H. R., and R. C. Seem. "Incidence-severity relationships for common maize rust on sweet corn." Phytopathology 80, no. 9 (1990): 842-849.

[10] Tamra A. Jackson, “Rust Diseases of Corn in Nebraska,” 2008. [Online]. Available: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=3816&context=extensionhist

[11] Sladojevic, Srdjan, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic. "Deep neural networks based recognition of plant diseases by leaf image classification." Computational intelligence and neuroscience 2016, no. 1 (2016): 3289801.

[12] Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." Frontiers in plant science 7 (2016): 215232.

[13] [Ahmadi, Parisa, Farrah Melissa Muharam, Khairulmazmi Ahmad, Shattri Mansor, and Idris Abu Seman. "Early detection of Ganoderma basal stem rot of oil palms using artificial neural network spectral analysis." Plant disease 101, no. 6 (2017): 1009-1016.

[14] Bhange, Manisha, and H. A. Hingoliwala. "Smart farming: Pomegranate disease detection using image processing." Procedia computer science 58 (2015): 280-288.

[15] Dixit, Aarju, and Sumit Nema. "Wheat leaf disease detection using machine learning method-a review." Int. J. Comput. Sci. Mob. Comput 7, no. 5 (2018): 124-129.

[16] S. Ghose, “Corn or Maize Leaf Disease Dataset,” 2020. [Online]. Available: https://www.kaggle.com/datasets/smaranjitghose/corn-or-maize-leaf-disease-dataset