
CGAN-AUGMENTED LIGHTWEIGHT CNN ARCHITECTURES WITH COORDINATE ATTENTION FOR MAIZE LEAF DISEASE RECOGNITION

Sani Saleh Saminu^{1,*}, Musa Idris Shehu¹, Yusuf Ibrahim¹, Sani Abdulwahab¹,
Yakubu Abubakar Hassan¹, Emejue Nichodemus Chibueze¹

¹Ahmadu Bello University, Zaria

Email: sssaleh@engineering.abu.edu.ng, Idris.ms@icloud.com, yibrahim@abu.edu.ng,
Saniabdulwahab499@gmail.com, yahassan@abu.edu.ng, emejue.chibueze@narict.gov.ng

*Corresponding author: Sani Saleh Saminu

DOI: <https://doi.org/10.67203/sp.ICESID.2026>

ABSTRACT

Deep learning-based plant disease recognition achieves strong benchmark results, but data scarcity and class imbalance consistently limit its reliability when deployed on real field imagery. This study comparatively evaluates four Coordinate Attention-enhanced lightweight CNNs, ShuffleNetV2, MobileNetV3-Small, GhostNet, and EfficientNet-B0, trained with Conditional GAN (cGAN) augmentation on the OSFmaize dataset across four disease categories and five data configurations. Generative quality assessment confirmed acceptable synthetic image quality, with best average FID of 198.57 at epoch 700 and best IS of 3.32 at epoch 1,000. In the primary Real + cGAN configuration, MobileNetV3-Small achieved the highest accuracy at 94.06% with 1.11M parameters, 0.061G FLOPs, 5.10ms inference, and 4.43MB storage, representing the best accuracy-efficiency trade-off among all four models. GhostNet recorded 80.07%, attributed directly to shortcut learning caused by training from scratch without pretrained weights, producing a 19.93 percentage point accuracy drop when real data was added to synthetic-only training. Traditional augmentation outperformed cGAN in all combined training scenarios. Ablation confirmed Coordinate Attention contributes up to 12.24 percentage points accuracy gain. These results provide practical guidance for architecture and augmentation selection in resource-constrained maize disease recognition systems.

Keywords: Maize leaf disease recognition; Conditional GAN; Lightweight CNN; Coordinate Attention; Data augmentation; MobileNetV3-Small

1. Introduction

Deep Deep learning has substantially changed how plant diseases are detected and classified, enabling automated diagnostic systems that work directly from digital leaf images without manual feature engineering (Li et al., 2021; Ouhami et al., 2021; Pardede et al., 2020). In precision agriculture, this shift matters practically: early disease identification reduces crop losses, cuts unnecessary pesticide use, and gives farmers actionable information before infections spread across a field. Maize (*Zea mays* L.) is one of the crops where this need is most acute. It is among the most widely grown cereals globally, with annual production exceeding 1.2 billion tonnes, yet it remains persistently vulnerable to foliar infections that can reduce yields by more than 50% in severe outbreaks (FAOSTAT, 2024; Solaimalai et al., 2020; USDA, 2024). Gray Leaf Spot (GLS), Northern Leaf Blight (NLB), and Northern Leaf Spot (NLS) are the primary culprits, and distinguishing between them accurately from leaf imagery is a non-trivial classification problem given their overlapping visual characteristics (Jasrotia et al., 2023; Langemeier et al., 2017).

Large-capacity architectures such as VGGNet (Simonyan & Zisserman, 2014), ResNet (He et al., 2016), and EfficientNet (Tan & Le, 2019) have demonstrated strong performance on plant disease benchmarks, but their practical utility depends on large, balanced training datasets. In real agricultural settings, assembling such datasets is consistently difficult. Seasonal variation, uneven disease incidence across fields, and inconsistent imaging conditions all produce small, class-imbalanced collections that push well-trained models toward overfitting (Li et al., 2021; Ouhami et al., 2021; Pardede et al., 2020).

Traditional augmentation methods, including rotation, flipping, color jittering, and random cropping, partially offset this by expanding dataset size, but they only introduce geometric and photometric variation. They do not produce new lesion morphologies, novel color gradients, or previously unseen severity patterns (Li et al., 2021). Models trained on traditionally augmented data can still fail to generalize when real field conditions differ from their training distribution.

Generative Adversarial Networks (GANs) address this limitation by learning the underlying distribution of real images and synthesizing new, semantically varied samples (Goodfellow et al., 2014). The Conditional GAN (cGAN; Mirza & Osindero, 2014), which conditions both the generator and discriminator on class labels, enables targeted synthesis of disease-specific leaf images that preserve class-relevant texture and color characteristics. Studies applying GAN-based augmentation to plant disease tasks, including tomato leaf detection (Wu et al., 2020) and corn disease identification (Hu et al., 2020), have reported measurable accuracy gains. However, GAN training for agricultural data introduces its own challenges. Training instability, mode collapse, and the tendency of models to memorize synthetic distributions rather than learn generalizable features are known failure modes that require careful evaluation (Saad et al., 2024; Wani et al., 2025). Systematic comparison of cGAN augmentation across multiple lightweight architectures on maize disease data, with explicit documentation of these failure conditions, remains limited in the literature.

Lightweight CNN architectures have attracted growing research interest for deployment on resource-constrained hardware, where large models are impractical (Chen et al., 2022; Le et al., 2024; Lu et al., 2023). ShuffleNetV2 (Ma et al., 2018), MobileNetV3, GhostNet, and EfficientNet-B0 each represent distinct trade-offs between parameter count, computational cost, and classification accuracy. Direct comparison of these models under identical training conditions provides useful guidance for practitioners who must select an architecture for edge deployment in agricultural settings. A recurring weakness across lightweight models is reduced sensitivity to spatially subtle lesions, since standard architectures apply attention uniformly across feature maps without discriminating between lesion and background regions (Bhujel et al., 2022). Coordinate Attention (Hou et al., 2021; Saminu et al., 2026) addresses this directly by encoding positional information along both horizontal and vertical spatial dimensions into the channel attention mechanism, allowing the model to localize fine-grained disease features without significant added computation. Prior work confirms attention-based improvements in lightweight plant disease classifiers (Bhujel et al., 2022; Ibrahim et al., 2025), but systematic multi-architecture comparisons under the same GAN-augmented training setup remain scarce.

This study addresses these gaps through a targeted problem-solution structure. Four challenges motivate the experimental design:

Limited dataset diversity. Small, imbalanced maize disease datasets cause overfitting and poor generalization across unseen field conditions. Traditional augmentation adds volume but not semantic variety. Solution: A cGAN trained for 1,000 epochs synthesizes class-conditioned maize leaf images for Healthy, GLS, NLB, and NLS categories, expanding both the size and visual diversity of the training set.

Inadequate spatial attention in lightweight models. Standard lightweight CNNs do not selectively attend to lesion regions, reducing their discriminative capacity when disease symptoms are subtle or unevenly distributed. Solution: Coordinate Attention is integrated at the final feature stage of each model, encoding directional spatial context into channel-wise feature recalibration without substantially increasing parameter count (Hou et al., 2021).

Absence of multi-architecture comparison under controlled conditions. Most existing studies evaluate a single model or compare architectures trained on different datasets and protocols, making it difficult to isolate the effect of architectural choice from training setup. Solution: Four lightweight CNNs ShuffleNetV2, MobileNetV3-Small, GhostNet, and EfficientNet-B0 are trained and evaluated under an identical protocol using a 70/15/15 stratified split with seed=42 across five data scenarios.

Deployment feasibility gap. Reporting accuracy alone is insufficient for field deployment decisions. Model size, inference speed, and parameter count must be evaluated alongside accuracy to determine practical suitability. Solution: All models are benchmarked on inference time, FLOPs, parameter count, and storage size to determine practical suitability for edge deployment.

This explicit problem-solution mapping ensures that each architectural and experimental decision addresses a specific, identified limitation rather than being arbitrarily assembled. The study represents a methodologically grounded comparison of cGAN augmentation and attention-enhanced

lightweight CNNs, not an additive combination of techniques.

To systematically evaluate this design, the study is guided by the following research questions:

- a) Which lightweight CNN architecture, enhanced with Coordinate Attention, achieves the best accuracy-efficiency trade-off when trained with cGAN-augmented maize disease data?
- b) Under what conditions does cGAN augmentation improve classification performance, and when does combining synthetic and real data degrade generalization?
- c) Does Coordinate Attention consistently improve classification performance across different lightweight architectures when applied under identical training conditions?

To answer these questions, this paper presents a comparative evaluation of four Coordinate Attention-enhanced lightweight CNNs trained across five data scenarios on the OSFmaize dataset. The experimental design incorporates cGAN-based augmentation with GAN quality evaluation using Fréchet Inception Distance (FID) and Inception Score (IS), and ablation studies isolating the Coordinate Attention contribution.

The main contributions of this work are as follows:

1. Multi-architecture comparison under controlled conditions: A systematic evaluation of ShuffleNetV2, MobileNetV3-Small, GhostNet, and EfficientNet-B0, each enhanced with Coordinate Attention, trained under an identical 70/15/15 stratified protocol across five data scenarios, enabling fair and reproducible architecture comparison.
2. cGAN augmentation analysis with failure documentation: An empirical characterization of cGAN augmentation behavior across five training scenarios, including explicit analysis of shortcut learning failure in architectures trained from scratch without pretrained weights, providing practical guidance for GAN-based augmentation in small agricultural datasets.
3. GAN quality evaluation across training checkpoints: Quantitative FID and IS evaluation at epochs 300, 500, 700, and 1,000, providing transparent evidence for the epoch selection decision used in downstream augmentation and documenting the trade-off between FID-optimal and IS-optimal checkpoints.

The remainder of this paper is organized as follows. Section 2 reviews related work on GAN-based augmentation and attention-enhanced lightweight CNNs for agricultural image classification. Section 3 describes the dataset preparation, cGAN training protocol, classifier architectures, and evaluation setup. Section 4 presents comparative results and ablation studies. Section 5 concludes with findings, limitations, and directions for future work.

2. Literature Review

A. GAN-Based Augmentation for Plant Disease Recognition

Data scarcity and class imbalance are persistent obstacles in agricultural image classification. Standard augmentation techniques such as rotation, flipping, and color jittering expand dataset volume but do not introduce new disease morphologies or lesion patterns. Generative Adversarial Networks, introduced by Goodfellow et al. (2014), address this by training a generator to produce synthetic images that a discriminator cannot reliably distinguish from real ones. The resulting synthetic samples add semantic variety that geometric transforms cannot provide.

For class-conditioned synthesis, the Conditional GAN (Mirza & Osindero, 2014) extends the original GAN by feeding class label information into both the generator and discriminator, allowing targeted generation of disease-specific images. This architecture is particularly suitable for small, imbalanced datasets where certain disease classes have insufficient training examples. Wu et al. (2020) applied DCGAN-based augmentation to tomato leaf disease classification and reported measurable accuracy improvements over models trained without generative augmentation. Mohana et al. (2022) reviewed GAN-based augmentation across multiple agricultural domains and confirmed that synthetic data consistently improves classifier performance when real data is limited, though they noted that image quality and training stability vary considerably across GAN architectures and dataset sizes.

In maize-specific tasks, Hu et al. (2020) demonstrated that GAN-augmented training data improved corn leaf disease identification accuracy by reducing inter-class confusion caused by visually

similar lesion patterns. Cap et al. (2020) introduced LeafGAN, a domain-adapted GAN for plant disease augmentation that achieved lower FID scores than standard cGAN on leaf datasets, confirming that domain-specific generator design improves synthetic image fidelity. Despite these advances, most existing studies evaluate a single GAN architecture on one downstream classifier. Systematic comparison of GAN augmentation behavior across multiple lightweight classifiers under identical training conditions remains absent, leaving practitioners without clear guidance on which architecture combinations generalize reliably.

A further concern with GAN-augmented training is the risk of shortcut learning, where models trained on synthetic-only data learn distributional shortcuts specific to the generator rather than true disease discriminators (Saad et al., 2024). This failure mode is particularly pronounced in architectures trained from scratch without pretrained weights, where the model has no prior visual knowledge to anchor feature learning. Saad et al. (2024) reviewed training challenges in GANs for biomedical image analysis and identified distribution mismatch between synthetic and real data as one of the most common causes of downstream classification degradation. Documenting these failure conditions explicitly is necessary for any study that proposes GAN augmentation as a practical tool.

B. Lightweight CNN Architectures for Agricultural Image Classification

Deploying disease recognition models in field conditions requires architectures that balance accuracy with computational efficiency. Standard deep networks such as VGGNet (Simonyan & Zisserman, 2014) and ResNet (He et al., 2016) deliver strong benchmark performance but carry parameter counts and inference times that are impractical on embedded or mobile hardware. Lightweight CNN architectures address this through structural innovations that reduce computation without sacrificing discriminative capacity.

ShuffleNetV2 (Ma et al., 2018) achieves efficiency through channel splitting and element-wise addition rather than computationally expensive group convolutions, resulting in fast inference and low memory usage. Chen et al. (2021) applied an improved ShuffleNetV2 to apple leaf disease recognition and confirmed that the architecture maintains competitive accuracy under resource constraints. MobileNetV3 builds on depthwise separable convolutions and incorporates hard-swish activations and squeeze-and-excitation blocks to improve accuracy per parameter. Abubakar et al. (2026) demonstrated the effectiveness of MobileNetV2 features compressed with PCA for maize disease recognition, achieving competitive accuracy with substantially reduced feature dimensionality, which reflects the broader utility of MobileNet-family architectures in agricultural classification tasks.

GhostNet reduces redundant feature map computation by generating additional feature maps from cheap linear operations applied to a small set of intrinsic features, cutting FLOPs without proportionally reducing representational capacity. EfficientNet-B0 (Tan & Le, 2019) applies compound scaling across network depth, width, and input resolution simultaneously, producing a model that achieves strong accuracy with fewer parameters than comparably performing architectures. Evangelista et al. (2023) evaluated EfficientNet V2 variants on maize disease classification and reported consistently high accuracy across disease categories, confirming EfficientNet-family models as reliable baselines for this task.

Lu et al. (2023) reviewed lightweight methods for plant disease identification and found that models with fewer than 5M parameters can match or approach the accuracy of larger networks when training data is sufficient and augmentation is applied effectively. Le et al. (2024) reached a similar conclusion using hybrid lightweight architectures for tomato disease detection, noting that inference speed and model size are as important as accuracy for real-world deployment. These findings collectively establish the four architectures used in this study as appropriate and well-motivated choices for comparative evaluation under cGAN-augmented training conditions.

C. ATTENTION MECHANISMS IN LIGHTWEIGHT PLANT DISEASE CLASSIFIERS

Lightweight CNNs process all spatial regions of a feature map with equal weight by default, which limits their ability to focus on diagnostically relevant areas such as lesion boundaries, color

transitions, or specific texture patterns. Attention mechanisms address this by learning to weight feature map regions according to their relevance to the classification task.

Squeeze-and-Excitation networks (Wang et al., 2020) introduced channel-wise attention by globally pooling spatial information and using a small fully connected network to recalibrate channel responses. Efficient Channel Attention (Wang et al., 2020) improved on this by replacing the fully connected layer with a lightweight 1D convolution, reducing parameter overhead while maintaining recalibration quality. Coordinate Attention (Hou et al., 2021) extends the spatial awareness of channel attention by decomposing global pooling into two directional pooling operations along the horizontal and vertical axes separately, encoding precise positional context into the attention weights. This directional encoding allows the model to localize spatially distributed features such as scattered lesions more accurately than standard channel attention, which discards spatial information entirely during pooling.

Bhujel et al. (2022) integrated a lightweight attention mechanism into a CNN for tomato leaf disease classification and reported consistent accuracy gains over the baseline without attention, particularly for visually similar disease classes. Ibrahim et al. (2025) combined depthwise separable convolutions with hybrid attention in EDANet for tomato disease recognition, confirming that attention improves both accuracy and feature discriminability in compact architectures. Rachman et al. (2024) applied attention-enhanced vision transformers to rice disease detection and found that cross-scale attention improved generalization to underrepresented disease classes. Jasrotia et al. (2023) evaluated CNN-based maize disease identification and noted that models lacking spatial attention mechanisms showed reduced performance on classes with fine-grained lesion textures, supporting the rationale for integrating CA into each architecture in this study.

Wang et al. (2020) confirmed through ablation that channel attention mechanisms consistently contribute to accuracy in lightweight CNNs across diverse visual classification tasks, with the contribution being largest when the model backbone has limited depth and therefore limited implicit spatial selectivity. This finding directly motivates applying CA at the final feature stage of each model in this study, where spatial selectivity has the greatest impact on downstream classification.

D. GAN Quality Evaluation: FID and Inception Score

Evaluating the quality of GAN-generated images requires metrics that go beyond visual inspection. Two quantitative measures are standard in the literature: Fréchet Inception Distance (FID) and Inception Score (IS). FID measures the distance between the distribution of real images and generated images in the feature space of a pretrained Inception-v3 network. Lower FID indicates that the synthetic distribution is closer to the real one, reflecting better perceptual realism and structural fidelity (Mohana et al., 2022). IS measures both image quality and class diversity within the generated set: a high IS means individual generated images are recognizable and the full set covers multiple distinct categories (Mohana et al., 2022).

Han et al. (2018) used FID to evaluate GAN-generated brain MRI images for data augmentation and found that lower FID correlated with better downstream classifier performance, establishing FID as a reliable proxy for augmentation utility in medical imaging. Cap et al. (2020) applied IS alongside FID to evaluate LeafGAN outputs on plant disease datasets, finding that IS scores above 2.0 were acceptable for domain-specific agricultural GANs where the total number of classes is small relative to ImageNet. Tang et al. (2019) similarly used FID to select optimal GAN training checkpoints for CT image augmentation, demonstrating that FID-based checkpoint selection is a practical and reproducible evaluation strategy.

In domain-specific datasets with few classes, IS values are expected to be considerably lower than those reported on ImageNet-scale benchmarks, since class diversity is inherently bounded by the number of categories in the dataset (Mohana et al., 2022). This distinction is important for interpreting IS results on the OSFmaize dataset, which contains only four classes. FID and IS together provide complementary evidence: FID captures distributional similarity to real data while IS captures per-image quality and intra-set diversity. When the two metrics favor different training checkpoints, as occurs in this study where FID is lowest at epoch 700 and IS is highest at epoch 1,000, the selection decision requires explicit justification rather than mechanical optimization of a single metric.

Existing work in agricultural GAN augmentation rarely reports both FID and IS across multiple training checkpoints, making it difficult to assess how generative image quality evolves during training and how checkpoint selection affects downstream classifier performance. This study addresses that gap by evaluating both metrics at epochs 300, 500, 700, and 1,000, providing a transparent basis for the augmentation epoch decision used across all five training scenarios.

3. PROPOSED FRAMEWORK

3.1 Overview

This section describes the dataset, preprocessing pipeline, cGAN implementation, classifier architectures, Coordinate Attention integration, and experimental configuration adopted for the comparative evaluation of four lightweight CNNs under cGAN-augmented training conditions for maize leaf disease recognition. The overall workflow integrates synthetic data generation using a Conditional GAN (cGAN) with four independently evaluated CA-enhanced lightweight classifiers to enable a controlled and fair multi-architecture comparison. The general framework is illustrated in Figure 3.

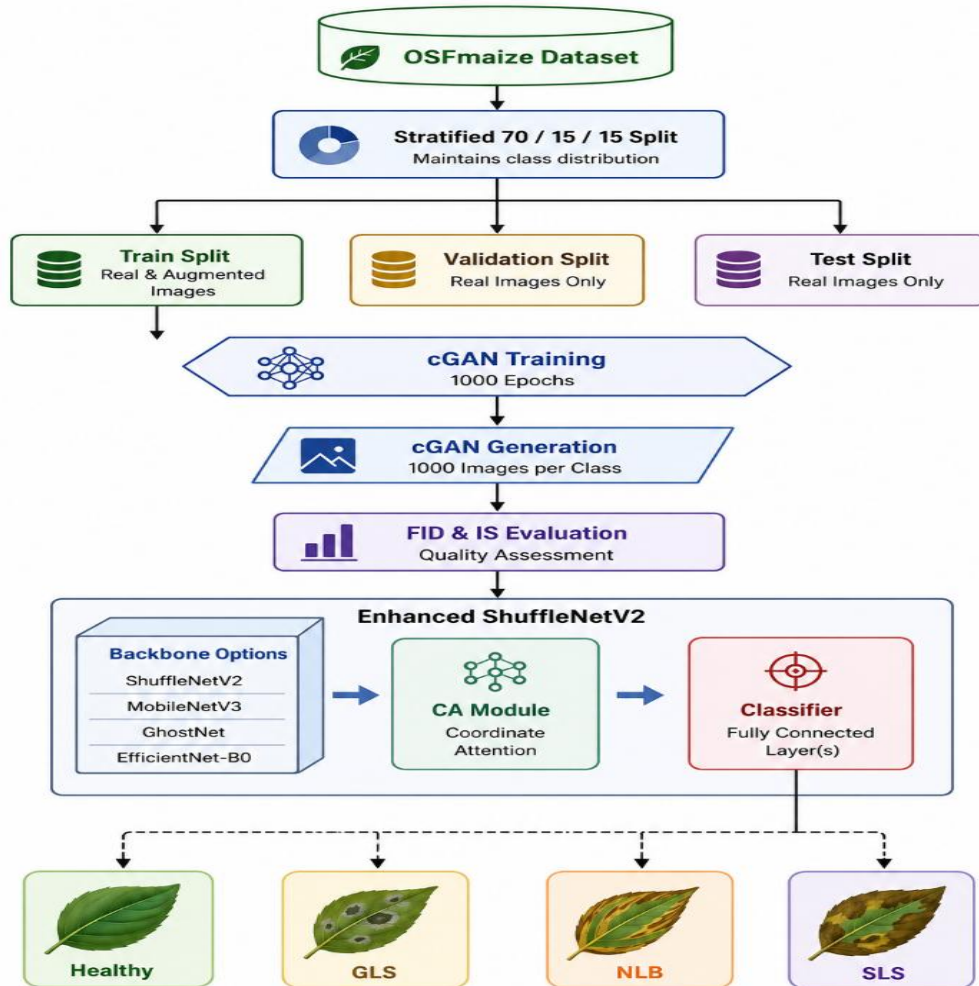


Figure 3: Developed DL Framework

The workflow trains a single cGAN on the training partition of the OSFmaize dataset and uses the generated images across all four classifier training scenarios. By holding the augmentation strategy, dataset split, evaluation protocol, and Coordinate Attention placement constant across all models, any observed performance differences can be attributed directly to architectural characteristics rather than experimental variation.

All experiments were implemented in Python 3.13 with PyTorch 2.10 on a local Ubuntu 24.04 workstation with the following hardware: HP Omen 16 laptop; Intel Core i9 14th-generation processor; 32 GB RAM; NVIDIA GeForce RTX 5070 GPU; CUDA 12.8. GAN training and classifier fine-tuning used the RTX 5070 GPU for acceleration. FID and IS evaluation was performed on CPU due to CUDA sm_120 architecture constraints with the installed PyTorch version, which does not affect metric correctness.

3.2 Dataset and Preprocessing

The dataset used in this study is the publicly available OSFmaize Leaf Disease dataset (<https://osf.io/s6ru5/>, accessed 17 September 2025). It comprises four classes representing distinct foliar conditions: Healthy, Gray Leaf Spot (GLS), Northern Leaf Blight (NLB), and Northern Leaf Spot (NLS), with a total of 1,902 real images across all classes.

After stratified splitting, the test set contains 286 images (15% of 1,902), with per-class counts proportional to the original class distribution.

Initial inspection revealed notable class imbalance, with Healthy and NLB containing substantially more samples than GLS and NLS. The images also exhibited visible heterogeneity in illumination, leaf orientation, and background clutter, reflecting the conditions of real-field acquisition. These factors can introduce bias into model learning if not addressed through principled dataset handling.

Data split and leakage prevention. The dataset was first partitioned into training (70%), validation (15%), and test (15%) subsets using stratified random sampling with a fixed random seed (seed = 42), prior to any synthetic generation. This three-way stratified split ensures class distribution is preserved across all partitions and that the test set provides an unbiased performance estimate independent of training decisions. The cGAN was trained exclusively on the training subset. Synthetic images were generated after cGAN training convergence and combined solely with the training partition to form the augmented training pool. The validation and test sets remained composed entirely of real, unseen images throughout all experiments. This procedure eliminates data leakage and ensures that all reported metrics reflect genuine generalization to unseen real-world data.

Each class received 1,000 cGAN-generated images. When combined with the real training partition, this produced a balanced augmented training corpus for the Real + cGAN scenario. All images were resized to $128 \times 128 \times 3$ pixels and normalized to the range $[-1, 1]$. No additional geometric augmentation was applied to the cGAN-generated images, since the generative process inherently introduces morphological diversity across samples.

The five experimental data configurations evaluated in this study are summarized in Table 3.1.

Table 3.1: Summary of the five experimental data configurations. Validation and test sets consist exclusively of real images in all configurations.

Configuration	Training Data	Val.	Test
Real Only	Real train split	Real only	Real only
cGAN Only	cGAN-generated images	Real only	Real only
Real + cGAN	Real train + cGAN images	Real only	Real only
Real + cGAN (no CA)	Real train + cGAN, no CA module	Real only	Real only
Real + TradAug	Real train + traditional augmentation	Real only	Real only

The 'Real + cGAN (no CA)' row represents an architectural ablation under the Real + cGAN data configuration, not an independent data scenario. The training data is identical to Real + cGAN; only the Coordinate Attention module is removed. It is included in this table for completeness but is discussed separately in Section 4.5.

3.3 Interpretation of FID and IS Metrics

Raw Frechet Inception Distance (FID) and Inception Score (IS) were used to quantify the perceptual realism and diversity of cGAN-generated images across training checkpoints. FID measures

the statistical distance between the feature distributions of real and synthetic images in the activation space of a pretrained Inception-V3 network. Lower FID indicates closer alignment with the real data distribution and higher visual fidelity. IS evaluates image quality and class diversity within the synthetic set alone; higher scores indicate more varied and semantically confident outputs.

FID is computed as:

$$FID = \left\| \mu_r - \mu_g \right\|^2 + \text{Tr} \left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}} \right) \quad (2.1)$$

where μ_r , Σ_r are the mean and covariance of real image features, and μ_g , Σ_g are the corresponding statistics for generated images extracted from the Inception-V3 pool3 layer.

IS is computed as:

$$IS = \exp \left(\mathbb{E}_{x \sim p_g} [D_{KL}(p(y|x) || p(y))] \right) \quad (2.2)$$

where $p(y|x)$ is the conditional class distribution predicted by Inception-V3 for a generated image x , and $p(y)$ is the marginal class distribution over all generated images. Higher IS reflects both high individual image confidence and diversity across the generated set.

Since Inception-V3 was pretrained on ImageNet rather than agricultural data, absolute FID and IS values are not directly comparable to benchmarks from natural image datasets. All FID and IS values reported here are used strictly for relative comparison across cGAN training checkpoints under identical evaluation conditions. Before computing FID and IS, all generated and real images at 128×128 resolution were resized to 299×299 pixels using bilinear interpolation to match the standard Inception-V3 input size, and pixel values were normalized to the range $[0, 1]$ prior to feature extraction.

3.4 Conditional GAN Implementation

The Conditional GAN (cGAN) (Miraz et al., 2024) generates class-specific maize disease images by conditioning both the generator and discriminator on class label information. The generator accepts a 100-dimensional random noise vector concatenated with a learned class label embedding, enabling synthesis of disease-specific leaf textures, colors, and lesion patterns. The discriminator receives real and generated image-label pairs and learns to distinguish genuine samples from synthetic ones, with the class label serving as an additional conditioning signal.

The adversarial objective is:

$$\begin{aligned} \mathcal{L}_{cGAN} = & \mathbb{E}_{(x, y)} [\log D(x | y)] \\ & + \mathbb{E}_{(z, y)} \left[\log \left(1 - D(G(z | y)) \right) \right] \end{aligned} \quad (2.3)$$

where x is a real image, y is the class label, z is a noise vector sampled from a prior distribution, G is the generator, and D is the discriminator. The generator minimizes this objective while the discriminator maximizes it, driving the generator toward producing class-consistent images that the discriminator cannot reliably reject.

The full cGAN training configuration is summarized in Table 3.2.

Table 3.2. cGAN training configuration.

Parameter	Setting	Description
Latent dimension	$z = 100$	Random noise vector input to generator
Batch size	64	Mini-batch per training iteration
Learning rate	0.0002	Adam optimizer ($b1 = 0.5$, $b2 = 0.999$)
Epochs	1,000	Total training epochs
Checkpoint epochs	300, 500, 700, 1,000	Epochs at which 1,000 images/class generated
Image size	$128 \times 128 \times 3$	Standardized RGB resolution
Optimizer	Adam	Used for both generator and discriminator
Hardware	NVIDIA RTX 5070 (HP Omen 16)	Ubuntu 24.04 local environment

3.5 Lightweight CNN Architectures with Coordinate Attention

Four lightweight CNN architectures were selected for comparative evaluation based on their distinct computational designs and demonstrated performance in resource-constrained classification tasks (Chen et al., 2022; Lu et al., 2023; Le et al., 2024). Each model was enhanced with a Coordinate Attention module inserted at the final convolutional feature stage, and each was independently trained and evaluated across the five data configurations. The classifier head was kept identical across all models: a fully connected layer (256 units) with ReLU activation, Dropout ($p = 0.4$) for regularization, and a final linear projection to four output classes.

3.4.1 ShuffleNetV2

ShuffleNetV2 (Ma et al., 2021) achieves computational efficiency through channel splitting and element-wise addition rather than group convolutions, which carry higher memory access costs. At each basic block, the input channels are split into two branches. One branch passes through unchanged while the other applies a sequence of 1x1 convolution, 3x3 depthwise convolution, and 1x1 convolution. The outputs are concatenated and shuffled to maintain cross-channel information flow. The variant used in this study is ShuffleNetV2 x1.0, which contains 1.62M parameters and operates at 0.153G FLOPs at 128 x 128 input resolution after CA integration. Pretrained ImageNet weights were loaded from local storage and used for full fine-tuning from epoch 1.

3.4.2 MobileNetV3-Small

MobileNetV3-Small [12] extends the MobileNet family with hard-swish activations, a redesigned last stage, and built-in Squeeze-and-Excitation (SE) blocks within its inverted residual structure. The hard-swish activation approximates the swish function with lower computational cost:

$$\text{h-swish}(x) = \frac{x \cdot \text{ReLU6}(x + 3)}{6} \quad (2.3)$$

The architecture is optimized through neural architecture search for minimal latency on mobile hardware. After CA integration and classifier replacement, the effective parameter count is 1.11M, operating at 0.061G FLOPs. Pretrained ImageNet weights were loaded locally and used for full fine-tuning.

3.4.3 GhostNet

GhostNet reduces feature map redundancy by generating additional feature maps from a small set of intrinsic features using cheap linear operations rather than computing all feature maps through standard convolutions. Given an input feature map X , the Ghost module produces:

$$Y' = \text{Conv}(X), \quad Y'' = \Phi(Y'), \quad Y = \text{concat}(Y', Y'') \quad (2.4)$$

where Y' is the set of intrinsic feature maps produced by a standard convolution, Φ represents the cheap linear transformation, and Y is the full output. GhostNet is implemented in this study using a native PyTorch implementation without the timm library or external pretrained weights, since no publicly available pretrained GhostNet weights matched the exact native configuration. The model is therefore trained from scratch across all experiments. After CA integration, the effective parameter count is 1.17M at 0.135G FLOPs.

3.4.4 EfficientNet-B0

EfficientNet-B0 (Tan et al., 2019) applies compound scaling, simultaneously increasing network depth, width, and input resolution according to a fixed scaling coefficient, to maximize accuracy per unit of computation. The compound scaling rule is:

$$\begin{aligned} d &= \alpha^\phi, & w &= \beta^\phi, & r &= \gamma^\phi \\ &= \gamma^\phi \quad \text{s.t.} & \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \end{aligned} \quad (2.5)$$

where d , w , and r denote depth, width, and resolution scaling factors respectively, ϕ is the compound coefficient, and α , β , γ are constants determined by grid search. After CA integration and classifier replacement, the effective parameter count is 4.49M at 0.416G FLOPs. Pretrained ImageNet weights were loaded locally and used for full fine-tuning.

3.6 Coordinate Attention Integration

Coordinate Attention (CA) (Hou et al., 2021) extends standard channel attention by preserving spatial location information along both the horizontal and vertical axes. CA decomposes global pooling into two directional operations:

$$z_h^c(h) = \frac{1}{W} \sum_{0 \leq i < W} x^c(h, i) \quad (2.6)$$

$$z_w^c(w) = \frac{1}{H} \sum_{0 \leq j < H} x^c(j, w) \quad (2.7)$$

where x^c is the c -th channel of the input feature map with height H and width W . The two directional descriptors are concatenated and passed through a shared transformation:

$$f = \delta(F_1([z^h, z^w])) \quad (2.8)$$

where δ is a non-linear activation and F_1 is a 1×1 convolution that reduces channel dimensionality by a reduction ratio r . The resulting embedding is split and projected back to generate attention weights along each spatial direction:

$$g^h = \sigma(F_h(f^h)), \quad g^w = \sigma(F_w(f^w)) \quad (2.9)$$

The final attended output is:

$$y^c(i, j) = x^c(i, j) \cdot g_c^h(i) \cdot g_c^w(j) \quad (2.10)$$

The CA module is inserted at the output of the final convolutional stage in each backbone, immediately before the global average pooling and classifier head, where spatial feature refinement has the greatest effect on downstream classification.

3.7 Experimental Configuration and Training Protocol

All four models were trained using cross-entropy loss with stochastic gradient descent (SGD), learning rate = 0.01, momentum = 0.9, and early stopping with patience = 7 epochs monitored on validation accuracy. Pretrained ImageNet weights were loaded locally for ShuffleNetV2, MobileNetV3-Small, and EfficientNet-B0. GhostNet was initialized randomly and trained from scratch. All experiments used a fixed seed of 42 for stratified splitting and reproducibility. Each model was evaluated on five data configurations as defined in Table 3.1, producing 20 main experimental runs in total. Ablation experiments removing the CA module were conducted for all four models on the Real + cGAN configuration to isolate the CA contribution.

All 20 main experimental runs and ablation experiments represent single training runs without repeated trials. Given the small dataset size (1,902 real images), results are sensitive to random initialization and split variance. Small percentage-point differences, particularly the 0.70 percentage point CA ablation result for ShuffleNetV2, should be interpreted with caution as they may not be distinguishable from random variation without repeated trials with variance reporting.

The Real + TradAug configuration applied the following augmentation pipeline to the real training images: random horizontal flip ($p = 0.5$), random vertical flip ($p = 0.5$), random rotation within ± 30 degrees, color jitter with brightness = 0.2, contrast = 0.2, saturation = 0.2, and hue = 0.1, and random resized crop to 128×128 with scale range [0.8, 1.0]. No synthetic images were used in this configuration.

3.8 Evaluation Criteria

Model performance was evaluated using accuracy, precision, recall, and F1-score computed on the real test set as in (Saminu et al., 2026; Bawa et al., 2026). Let TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives respectively:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.11)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.12)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.13)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (2.14)$$

Weighted-average metrics were used across all four classes to account for class distribution differences in the test set. GAN image quality was assessed using FID and IS as described in Section 3.3. Inference time was measured as the average time per image on the RTX 5070 GPU across the full test set.

4. RESULTS AND ANALYSIS

This section presents and discusses the experimental results across all five data configurations for the four CA-enhanced lightweight CNN architectures. Results are organized as follows: cGAN training and image quality evaluation, main classification results, ablation analysis of Coordinate Attention, augmentation strategy comparison, convergence analysis, per-class classification analysis, and efficiency benchmarking.

4.1 cGAN Training Quality Evaluation

Figure 4.1 shows the cGAN training loss curve over 1,000 epochs. The discriminator loss stabilized near zero early in training and remained consistently low throughout, indicating that the discriminator learned a reliable decision boundary quickly. The generator loss increased gradually over training epochs, reflecting the expected adversarial dynamic where a stronger discriminator forces the generator to produce progressively harder-to-reject samples. No mode collapse was observed across the 1,000-epoch run.

Table 4.1 summarizes the FID and IS scores computed at epochs 300, 500, 700, and 1,000 across all four disease classes.

Table 4.1: FID and IS evaluation across cGAN training checkpoints

Epoch	Health FID	GLS FID	NLB FID	NLS FID	Avg FID	Combined IS
300	171.23	241.84	257.66	225.24	223.99	2.98
500	144.58	223.26	245.76	210.57	206.04	3.06
700	142.55	213.53	233.48	204.73	198.57	3.28
1,000	129.83	228.31	238.33	207.66	201.03	3.32

FID improved consistently from epoch 300 (avg 223.99) to epoch 700 (avg 198.57), then rose slightly at epoch 1,000 (avg 201.03). IS increased monotonically across all checkpoints, reaching its highest value of 3.32 at epoch 1,000. This creates a trade-off: epoch 700 produces synthetic images with better distributional similarity to real data, while epoch 1,000 produces images with higher individual quality and intra-set diversity.

The Healthy class consistently achieved the lowest FID across all checkpoints, reaching 129.83 at epoch 1,000. This is expected given that healthy leaf images are visually more uniform than diseased ones. GLS, NLB, and NLS showed higher FID values throughout, reflecting the greater visual

complexity and inter-class similarity of disease lesion patterns. All IS values above 2.0 across all epochs are consistent with the reference threshold reported by Cap et al. [21] for domain-specific agricultural GANs, though this threshold was derived from a different dataset and serves here as a heuristic reference point rather than a universally established benchmark.

Epoch 1,000 was selected for downstream augmentation experiments because it produced the highest IS (3.32), indicating superior image diversity, and its FID (201.03) remained comparable to epoch 700 (198.57), with a difference of only 2.46 points. Given the small FID gap and the meaningful IS gain, epoch 1,000 represents the better overall trade-off for augmenting a small, imbalanced dataset where intra-class diversity directly affects classifier generalization.

It should be noted that the downstream classifier accuracy using epoch-700-generated images was not separately evaluated. The epoch 1,000 selection is based on the FID/IS trade-off analysis above rather than a direct empirical comparison of downstream classification performance between the two checkpoints.

4.2 Main Classification Results: Real + cGAN ep1000

Table 4.2 presents the main classification results for all four models trained on the Real + cGAN ep1000 configuration. Figure 4.2 shows the model comparison across accuracy, F1-score, inference time, and parameter count.

Table 4.2: Main classification results: Real + cGAN ep1000 with Coordinate Attention

Model	Acc (%)	Prec (%)	Rec (%)	F1 (%)	Params (M)	FLOPs (G)	Inf (ms)	Size (MB)
MobileNetV3-Small	94.0	94.6	94.0	93.9	1.1	0.06	5.10	4.43
ShuffleNetV2	93.3	93.4	93.3	93.3	1.6	0.15	9.17	6.46
EfficientNet-B0	88.8	89.1	88.8	88.8	4.4	0.41	30.8	17.9
GhostNet	80.0	80.9	80.0	79.8	1.1	0.13	15.8	4.66

MobileNetV3-Small achieved the best overall result at 94.06% accuracy and 93.99% F1-score, with the smallest parameter count (1.11M), lowest FLOPs (0.061G), fastest inference (5.10ms), and second-smallest model size (4.43MB). ShuffleNetV2 followed closely at 93.36% accuracy with a 9.17ms inference time and 1.62M parameters. The 0.70 percentage point accuracy gap between the two is small, but MobileNetV3-Small's inference speed advantage (5.10ms vs 9.17ms) is practically significant for real-time field applications.

EfficientNet-B0 achieved 88.81% accuracy despite having the largest parameter count (4.49M), highest FLOPs (0.416G), and slowest inference time (30.83ms) among the four models. Its larger capacity did not translate to better performance under cGAN-augmented training on this dataset size, suggesting that the additional model complexity was not justified by the available training data. GhostNet recorded the lowest accuracy at 80.07%, a result directly explained by its training-from-scratch condition, discussed further in Section 4.4.

4.3 Results Across All Five Data Scenarios

Table 4.3 presents the complete accuracy results across all five configurations for all four models.

Table 4.3: Classification accuracy (%) across all five data configurations

Model	Real Only	cGAN Only	Real cGAN	+ Real + cGAN (no CA)	Real TradAug	+
MobileNetV3-Small	93.01	99.83	94.06	81.82	98.60	
ShuffleNetV2	32.87	99.67	93.36	94.06	98.25	
EfficientNet-B0	95.45	99.83	88.81	85.66	97.20	
GhostNet	91.61	100.00	80.07	74.83	96.15	

Several patterns emerge across scenarios. All four models achieved near-perfect or perfect accuracy under cGAN-only training (99.67% to 100.00%). This result does not reflect genuine generalization. When trained on synthetic data alone and tested on real images, the high accuracy indicates the models learned the synthetic data distribution effectively, but the real test set happened to be classifiable from features the generator had already encoded. The collapse in performance when real data is added in the Real + cGAN scenario, particularly for GhostNet (100.00% to 80.07%), confirms that synthetic-only accuracy does not predict real-world generalization.

Traditional augmentation outperformed cGAN augmentation in the Real + Aug comparison for all four models. MobileNetV3-Small reached 98.60% with TradAug versus 94.06% with cGAN. ShuffleNetV2 reached 98.25% with TradAug versus 93.36% with cGAN. Geometric and photometric augmentation of real images produces more in-distribution training samples than GAN-synthesized data, which can introduce distributional artifacts that partially conflict with real image statistics (Wu et al., 2020).

ShuffleNetV2 recorded only 32.87% accuracy in the Real Only scenario. The Real Only training set contains only 1,331 images split across four classes after stratified partitioning. ShuffleNetV2's channel-split architecture appears particularly sensitive to small dataset size without augmentation. The convergence analysis for the Real Only run indicates the model stopped early, suggesting it did not have sufficient data to learn stable class boundaries before convergence criteria were triggered; however, no dedicated training curve for this specific scenario is reported here, and this explanation should be treated as inferred rather than directly observed. The addition of 4,000 cGAN-generated images in the Real + cGAN scenario resolved this, confirming that ShuffleNetV2 requires a minimum data volume to function effectively.

4.4 GhostNet Shortcut Learning Analysis

GhostNet's result in the cGAN-only scenario warrants specific attention. A perfect 100.00% accuracy on the real test set when trained exclusively on synthetic data is not a sign of a strong model. It indicates shortcut learning: the model memorized the cGAN output distribution so effectively that the features it learned aligned with the real test set by coincidence rather than genuine disease discrimination.

When real images were added in the Real + cGAN configuration, accuracy dropped to 80.07%, a 19.93 percentage point decline. Adding real images introduced a second, conflicting distribution that the model could not reconcile without pretrained weights. The other three models, all initialized from ImageNet weights, showed no such collapse. MobileNetV3-Small, ShuffleNetV2, and EfficientNet-B0 all performed stably when moving from cGAN-only to Real + cGAN training. The contrast confirms that pretrained initialization is a necessary condition for stable generalization under combined real and synthetic training on small agricultural datasets.

The GhostNet confusion matrix (Figure 4.3b) shows the clearest evidence of this failure: 31 of 79 GLS samples were misclassified as NLB, and 12 of 75 NLB samples were misclassified as GLS. These two classes share streak-like lesion textures, and without pretrained low-level feature representations, GhostNet could not learn sufficiently discriminative boundaries between them from the

combined dataset.

A methodological limitation of this comparison should be acknowledged: GhostNet was trained from scratch due to the absence of publicly available pretrained weights compatible with the native PyTorch implementation used, while ShuffleNetV2, MobileNetV3-Small, and EfficientNet-B0 were initialized from ImageNet weights. This pretraining difference is an uncontrolled variable in the multi-architecture comparison. GhostNet's lower performance should therefore be interpreted as initialization-dependent rather than a property of the Ghost module architecture in isolation. A fairer architectural comparison would require pretrained GhostNet weights, which are available via timm for some configurations and should be explored in future work.

4.5 Coordinate Attention Ablation

Table 4.4 presents the ablation results comparing each model with and without the Coordinate Attention module on the Real + cGAN ep1000 configuration.

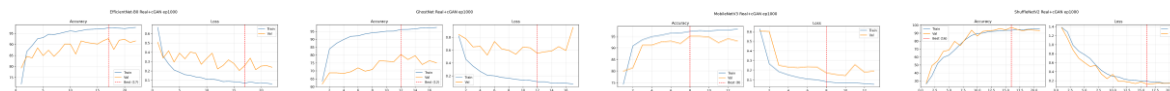
Table 4.4: Coordinate Attention ablation results: Real + cGAN ep1000

Model	With CA (%)	Without CA (%)	CA Gain (pp)
MobileNetV3-Small	94.06	81.82	+12.24
GhostNet	80.07	74.83	+5.24
EfficientNet-B0	88.81	85.66	+3.15
ShuffleNetV2	93.36	94.06	-0.70

Overall, Coordinate Attention produced consistent accuracy gains for three of the four models. The largest gain was in MobileNetV3-Small, where removing CA caused a 12.24 percentage point drop from 94.06% to 81.82%. MobileNetV3-Small's built-in SE blocks operate on global channel statistics and discard spatial position information. CA's directional pooling recovers this spatial context, allowing the model to attend to localized lesion regions rather than treating all spatial positions equally. The combination of SE and CA in MobileNetV3-Small thus provides complementary rather than redundant attention.

GhostNet gained 5.24 percentage points with CA, and EfficientNet-B0 gained 3.15 points. Both results confirm that directional spatial attention improves lesion localization across architectures with different internal structures. ShuffleNetV2 showed a marginal 0.70 percentage point decrease with CA compared to without. The channel-shuffle operation in ShuffleNetV2 already promotes cross-channel information mixing, which may partially substitute for the recalibration that CA provides in architectures without such mixing

4.6 Convergence Analysis

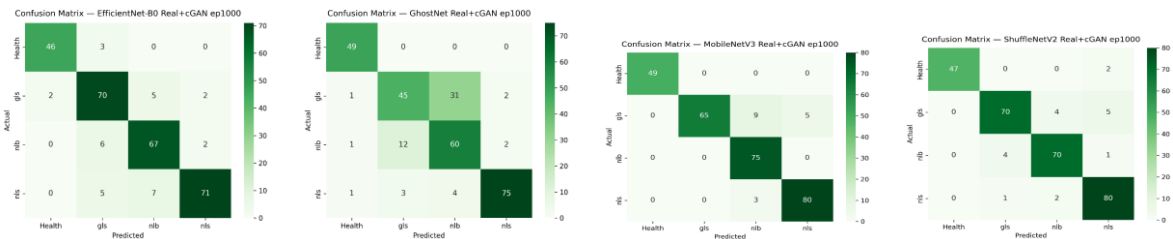


Figures 4.5 through 4.8 show the training and validation accuracy and loss curves for each model under the Real + cGAN ep1000 configuration. MobileNetV3-Small (Figure 4.7) converged fastest, reaching its best validation accuracy of 95% at epoch 8 before early stopping triggered at epoch 13. Both train and validation loss decreased smoothly and in close alignment, with no sign of overfitting.

ShuffleNetV2 (Figure 4.8) showed a longer convergence trajectory, reaching its best validation accuracy at epoch 16 out of 21 total epochs. The training and validation curves tracked closely throughout, with both loss curves descending steadily. EfficientNet-B0 (Figure 4.5) reached its best validation accuracy at epoch 17, with training accuracy exceeding 97% while validation accuracy stabilized around 91-93%. The visible gap between training and validation accuracy from epoch 10 onward suggests mild overfitting, attributable to EfficientNet-B0's larger capacity relative to the effective training set size.

GhostNet (Figure 4.6) displayed the most unstable convergence pattern. Validation accuracy peaked at 80% at epoch 12, after which validation loss began rising sharply while training loss continued to decrease. This divergence between training and validation loss is the signature of overfitting, directly attributable to the absence of pretrained weights. Without ImageNet initialization, GhostNet must learn feature representations entirely from the training data, and the distribution mismatch between real and synthetic images creates optimization instability that pretrained models avoid.

4.7 Per-Class Classification Analysis



Figures 4.3a through 4.3d show the confusion matrices for all four models on the Real + cGAN ep1000 test set. MobileNetV3-Small (Figure 4.3c) produced the cleanest classification pattern. It correctly classified all 49 Healthy samples, all 75 NLB samples, and 80 of 83 NLS samples. The only notable errors were 9 GLS samples misclassified as NLB and 5 GLS samples misclassified as NLS, out of 79 total GLS samples. GLS was the only class where the model showed consistent confusion, expected given that GLS lesion textures share angular streak characteristics with NLB at certain disease stages.

ShuffleNetV2 (Figure 4.3d) correctly classified 47 of 49 Healthy samples, 70 of 79 GLS samples, 70 of 75 NLB samples, and 80 of 83 NLS samples. Errors were distributed relatively evenly across classes, with no single class dominating the misclassification pattern. EfficientNet-B0 (Figure 4.3a) correctly classified 46 of 49 Healthy samples and 71 of 83 NLS samples. Its GLS-NLB misclassifications (11 total) are comparable to MobileNetV3-Small (14 total), and its lower overall accuracy relative to the other pretrained models is distributed across multiple class boundaries rather than concentrated at a single pair.

GhostNet (Figure 4.3b) showed the highest concentration of errors at the GLS-NLB boundary. Thirty-one of 79 GLS samples were predicted as NLB, representing a 39.2% error rate for that class. This single failure accounts for a large portion of GhostNet's overall accuracy deficit and confirms that the GLS-NLB boundary is the primary point of failure for models without stable pretrained feature representations.

4.8 Efficiency Analysis

Table 4.5 summarizes the efficiency profile of all four models under the Real + cGAN ep1000 configuration

Table 4.5: Model efficiency comparison: Real + cGAN ep1000 with Coordinate Attention.

Model	Params (M)	FLOPs (G)	Inf (ms)	Size (MB)	Acc (%)
MobileNetV3-Small	1.11	0.061	5.10	4.43	94.06
GhostNet	1.17	0.135	15.86	4.66	80.07
ShuffleNetV2	1.62	0.153	9.17	6.46	93.36
EfficientNet-B0	4.49	0.416	30.83	17.96	88.81

MobileNetV3-Small achieves the best accuracy-efficiency trade-off across all four metrics simultaneously. Its FLOPs of 0.061G are 2.2 times lower than GhostNet (0.135G), 2.5 times lower than ShuffleNetV2 (0.153G), and 6.8 times lower than EfficientNet-B0 (0.416G). Its inference time of 5.10ms is the fastest in the group and its 4.43MB storage footprint is the smallest. EfficientNet-B0's 4.49M parameters and 17.96MB storage size make it the least suitable for edge deployment, and its 30.83ms inference time is six times slower than MobileNetV3-Small while delivering 5.25 percentage points lower accuracy in this scenario.

5. CONCLUSION

This study comparatively evaluated four Coordinate Attention-enhanced lightweight CNNs under cGAN augmentation for maize leaf disease recognition. MobileNetV3-Small achieved the best result at 94.06% accuracy with 1.11M parameters, 0.061G FLOPs, and 5.10ms inference, making it the most suitable architecture for resource-constrained deployment. GhostNet's performance collapsed from 100.00% under synthetic-only training to 80.07% when real data was added, confirming that pretrained initialization is necessary for stable generalization under combined real and synthetic training. Traditional augmentation outperformed cGAN in all combined scenarios, establishing that synthetic generation is most valuable when real data is severely limited. Ablation confirmed Coordinate Attention contributes consistent gains across three of four architectures, with the largest improvement of 12.24 percentage points in MobileNetV3-Small, where CA's directional spatial encoding complements the model's built-in global channel attention. The primary finding is that no single architecture dominates across all criteria: MobileNetV3-Small leads on accuracy and efficiency, while ShuffleNetV2 offers competitive accuracy at moderate cost. Future work will evaluate WGAN-GP and StyleGAN2-ADA as alternative augmentation strategies, test robustness on UAV-captured field imagery, and explore structured pruning of MobileNetV3-Small for low-power edge deployment.

Acknowledgements

The authors thank the Department of Computer Engineering, Ahmadu Bello University, Zaria, Nigeria.

REFERENCES

- Solaimalai, A., Anantharaju, P., Irulandi, S., & Theradimani, M. (2020). *Maize crop: Improvement, production, protection and post-harvest technology*. CRC Press.
- FAOSTAT. (2024). *Crops processed*. Food and Agriculture Organization. <https://www.fao.org/faostat/>
- USDA. (2024). *Corn and other feed grains*. Economic Research Service. <https://www.ers.usda.gov/topics/crops/corn-and-other-feed-grains/>
- Langemeier, C. B., Robertson, A. E., Wang, D., Jackson-Ziems, T. A., & Kruger, G. R. (2017). Factors affecting the development and severity of Goss's wilt of corn. *Plant Disease*, 101(1), 54–61.
- Sharma, S., Kumar, V., & Sood, S. (2023). Pest detection using random forest. *Proceedings of the International Conference on Innovative Computing and Communications and Automation of India (ICICAT)*, 1–8.

- Khirade, S. D., & Patil, A. B. (2015). Plant disease detection using image processing. *Proceedings of the International Conference on Computing Communication Control and Automation (ICCCCA)*, 768–771.
- Pardede, H. F., Suryawati, E., Zilvan, V., Ramdan, A., Heryana, A., & Kusumo, R. B. S. (2020). Machine learning based plant diseases detection: A review. *Proceedings of the International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, 212–217.
- Li, L., Zhang, S., & Wang, B. (2021). Plant disease detection and classification by deep learning: A review. *IEEE Access*, 9, 56683–56698.
- Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., & Canals, R. (2021). Computer vision, IoT and data fusion for crop disease detection using machine learning. *Remote Sensing*, 13(13), Article 2486.
- Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv. <https://arxiv.org/abs/1409.1556>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 6105–6114.
- Paymode, P., & Malode, V. (2022). Transfer learning for multi-crop leaf disease classification using VGG. *Artificial Intelligence in Agriculture*, 6, 23–33.
- Ullah, N., Khan, J. A., Alharbi, L. A., Raza, A., Khan, W., & Ahmad, I. (2023). An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model. *Frontiers in Plant Science*, 14, Article 1079366.
- Tiwari, V., Joshi, R. C., & Dutta, M. K. (2021). Dense convolutional neural network based multiclass plant disease detection and classification using leaf images. *Ecological Informatics*, 63, Article 101289.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2672–2680.
- Mirza, M., & Osindero, S. (2014). *Conditional generative adversarial nets*. arXiv. <https://arxiv.org/abs/1411.1784>
- Odena, A., Olah, C., & Shlens, J. (2017). Conditional image synthesis with auxiliary classifier GANs. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2642–2651.
- Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 214–223.
- Han, C., Murao, K., Noguchi, T., Kawata, Y., Uchiyama, F., Rundo, L., Nakayama, H., & Hayashi, H. (2018). GAN-based synthetic brain MR image generation. *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, 734–738.
- Cap, Q. H., Uga, H., Kagiwada, S., & Iyatomi, H. (2020). LeafGAN: An effective data augmentation method for practical plant disease diagnosis. arXiv. <https://arxiv.org/abs/2002.10100>
- Tang, Y., Tang, Y., Zhu, Y., Summers, R. M., & Xiao, J. (2019). CT-realistic data augmentation using generative adversarial network for robust lymph node segmentation. *Proceedings of SPIE Medical Imaging*, Article 109503W.
- Chen, Y., Wen, C., & Wen, C. (2022). DFCA Net: A novel lightweight convolutional neural network model for corn disease identification. *Agriculture*, 12(12), Article 2047.
- Lu, J., Tan, L., & Jiang, H. (2023). Lightweight method for plant disease identification using deep learning. *Intelligent Automation and Soft Computing*, 37, 525–544.
- Le, T., Nguyen, T., Nguyen, B., & Tran, T. (2024). Tomato disease detection with lightweight recurrent and convolutional neural network models. *Frontiers in Sustainable Food Systems*, 5, Article 807011.

- Yang, X., Gao, Z., Li, H., Li, Z., Zhao, F., Wu, X., & Zhu, Y. (2024). A convolutional neural network and transformer framework for leaf disease and severity identification. *Scientific Reports*, 14, Article 3023.
- Wu, Q., Chen, Y., & Meng, J. (2020). DCGAN-based data augmentation for tomato leaf disease identification. *IEEE Access*, 8, 98727–98739.
- Ahmed, M., & Ahmed, A. (2023). Palm tree disease detection and classification using residual network and transfer learning. *PLOS ONE*, 18(3), Article e0283250.
- Mohana, R. M., Reddy, C. V. R., Madhavi, K., Supreethi, K. P., Kora, P., Thankachan, B., & Pavithra, L. (2022). *Generative adversarial networks and its applications in the biomedical image analysis and remote sensing image analysis: A review*. arXiv. <https://arxiv.org/abs/2204.04707>
- Ma, Y., Zhang, Z., Kang, J., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Computers and Electronics in Agriculture*, 187, Article 106314.
- Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13713–13722.
- Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11534–11542.
- Ma, N., Zhang, X., Zheng, H.-T., & Sun, J. (2018). ShuffleNet V2: Practical guidelines for efficient CNN architecture design. *Proceedings of the European Conference on Computer Vision (ECCV)*, 116–131.
- Chen, C., Zhu, X., & Liu, C. (2021). Apple leaf disease recognition based on improved ShuffleNet V2. *Proceedings of the International Conference on Culture-oriented Science and Technology (CTISC)*, 276–282.
- Bari, B. S., Islam, M. N., Rashid, M., Hasan, M. J., Razman, M. A. M., Musa, R. M., Ab Nasir, A. F., & Majeed, A. P. P. A. (2021). A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN. *PeerJ Computer Science*, 7, Article e432.
- Bhujel, A., Kim, N.-E., Arulmozhi, E., Basak, J. K., & Kim, H.-T. (2022). A lightweight attention-based convolutional neural network for tomato leaf disease classification. *Agriculture*, 12(2), Article 228.
- Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6232–6251.
- Murray, N., & Perronnin, F. (2014). Generalized max pooling. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2473–2480.
- Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., & Shen, H. (2020). Single image super-resolution via a holistic attention network. *Proceedings of the European Conference on Computer Vision (ECCV)*, 191–207.
- Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023). Efficient multi-scale attention module with cross-spatial learning. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.
- Hu, R., Hou, J., Liu, J., & Liu, L. (2020). Corn leaf disease identification based on transfer learning and data augmentation. *Proceedings of the International Conference on Computer Science and Systems Engineering (ICCSSE)*, 1–6.
- Evangelista, I., Flores, N., Zafra, R., & Culaba, I. (2023). Zea mays (corn) multi-disease classification using EfficientNet V2 variants. *Proceedings of the IEEE Region 10 Symposium (TENSYP)*, 1–5.
- Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, 61, Article 101182.

- Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., & Keutzer, K. (2014). *DenseNet: Implementing efficient ConvNet descriptor pyramids*. arXiv. <https://arxiv.org/abs/1404.1869>
- Zhao, Y., Chen, X., Guo, M., Wang, H., & Wang, H. (2021). SEV-Net: Residual network embedded with attention mechanism for plant disease severity detection. *Concurrency and Computation: Practice and Experience*, 33(20), Article e6161.
- Yang, W., Zheng, Y., Wu, Y., & Li, Y. (2022). Few-shot unsupervised domain adaptation via meta learning. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.
- Singh, S. R., & Sanodiya, R. K. (2023). A zero-shot transfer learning framework for plant leaf disease classification. *IEEE Access*, 11, 143861–143880.
- Wu, X., Zhan, C., Lai, Y.-K., Cheng, M.-M., & Yang, J. (2023). Unsupervised domain adaptation for plant disease recognition in the wild. *Plant Phenomics*, 5, Article 0010.
- Li, Y., & Chao, X. (2021). Semi-supervised few-shot learning approach for plant diseases recognition. *Plant Methods*, 17(1), Article 1.
- Ibrahim, Y., Momoh, M. O., Shobowale, K. O., Abubakar, Z. M., & Yahaya, B. (2025). EDANet: A novel architecture combining depthwise separable convolutions and hybrid attention for efficient tomato disease recognition. *Journal of Computing Theories and Applications*, 3(2), 160–170.
- Rachman, R. K., Setiadi, D. R. I. M., Susanto, A., Nugroho, K., & Islam, H. M. M. (2024). Enhanced vision transformer and transfer learning approach to improve rice disease recognition. *Journal of Computing Theories and Applications*, 1(4), 446–460.
- Jasrotia, S., Yadav, J., Rajpal, N., Arora, M., & Chaudhary, J. (2023). Convolutional neural network based maize plant disease identification. *Procedia Computer Science*, 218, 1712–1721.
- Saad, M. M., O'Reilly, R., & Rehmani, M. H. (2024). A survey on training challenges in generative adversarial networks for biomedical image analysis. *Artificial Intelligence Review*, 57, Article 19. <https://doi.org/10.1007/s10462-023-10624-y>
- Wani, M. A., Sultan, B., Ali, S., & Sofi, M. A. (2025). Generative adversarial networks in image steganography. In *Advances in deep learning: Vol. 2. Studies in big data* (Vol. 12). Springer. https://doi.org/10.1007/978-981-96-3498-9_5
- Abubakar, M., Ibrahim, Y., Ajayi, O. O., & Saminu, S. S. (2026). A lightweight maize leaf disease recognition using PCA-compressed MobileNetV2 features and RBF-SVM. *Journal of Computing Theories and Applications*, 3(3), 334–348.
- Bawa, M. S., Yusuf, S. M., & Saminu, S. S. (2026). MSA-TCN: Robust urban suspicious sound detection using multi-scale temporal convolutions and dual attention. *Journal of Future Artificial Intelligence and Technologies*, 3(1), 36–52.
- Saminu, S. S., Ibrahim, Y., Haruna, Z., & Umar, A. (2026). Enhanced cross-crop plant disease recognition model via domain-specific alignment network and elastic-mixup. *International Journal of Software Engineering and Computer Systems*, 12(1), 39–49.