A Comprehensive Image Dataset of Fruit and Leaf Diseases Across Six Horticultural Crops for Deep Learning Applications

anand Kumar jain

2023rcp9026@mnit.ac.in

MNIT: Malaviya National Institute of Technology Jaipur https://orcid.org/0009-0004-7184-8479

Neeta Nain

Research Article

Keywords: Apple, Banana, Citrus, Guava, Mango, Papaya, ABCGMP dataset, Fruit, Leaf, CNN

Posted Date: March 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9207865/v1

A Comprehensive Image Dataset of Fruit and Leaf Diseases Across Six Horticultural Crops for Deep Learning Applications

Anand Kumar Jain and Neeta Nain

Department of CSE MNIT, Jaipur (India) 2023rcp9026@mnit.ac.in, nnain.cse@mnit.ac.in Corresponding Author: Anand Kumar Jain Corresponding Author Email: 2023rcp9026@mnit.ac.in

Abstract. Accurate and timely identification of plant diseases is essential for improving crop productivity and ensuring sustainable agricultural practices. This paper presents a comprehensive image dataset of fruit and leaf diseases covering six economically important horticultural crops: Apple, Banana, Citrus, Guava, Mango, and Papaya. The dataset comprises high-quality RGB images representing both healthy and diseased samples, with disease symptoms including spots, lesions, discoloration, blight, rot, and fungal and bacterial infections captured under diverse real-world conditions. Variations in illumination, background complexity, viewing angles, growth stages, and symptom severity are intentionally included to enhance the robustness and generalizability of learning models developed using this data.
The dataset is structured in a class-wise manner and preprocessed to support direct integration with deep learning frameworks. It is extensively used to train, validate, and evaluate deep learning–based plant disease classification models, enabling automatic feature learning from raw images without manual intervention. Experimental usage demonstrates that the dataset is well suited for convolutional neural networks and attention-based architectures, facilitating effective discrimination between multiple disease categories across different crops and plant organs. By providing a unified multi-crop, multi-disease benchmark, this dataset aims to accelerate research in automated crop disease diagnosis, precision agriculture, and intelligent decision-support systems for sustainable farming.

Keywords: Apple · Banana · Citrus · Guava · Mango · Papaya · ABCGMP dataset · Fruit · Leaf · CNN.

1 Introduction

Accurate and timely detection of plant diseases [1] is essential for improving crop productivity, reducing economic losses, and supporting sustainable agricultural practices. Although deep learning techniques have shown strong potential for automated disease diagnosis, their effectiveness is largely constrained by the quality and realism of available datasets. Most existing plant disease datasets are either crop-specific, limited to a single plant organ, or collected under controlled conditions, which restricts model generalization to real-field environments.

To address these limitations, this work presents the **ABCGMP Fruit and Leaf Disease Dataset**, a comprehensive multi-crop dataset covering apple, banana, citrus, guava, mango, and papaya. The dataset includes both fruit and leaf images captured under diverse field conditions, reflecting real-world variability in lighting, background, and disease severity. By enabling multi-crop and multi-organ disease analysis, the ABCGMP dataset provides a robust benchmark for developing and evaluating generalized deep learning models for practical plant disease diagnosis.

1.1 Apple Crop and Disease Overview

Fig. 1. Apple disease and Healthy sample of ABCGMP datasets [23].

Apple (*Malus domestica*) is a major temperate fruit crop of high nutritional and economic importance. Optimal apple production requires cool winters and mild summers, with an ideal temperature range of 18–24 °C and a winter chilling requirement of 800–1500 hours below 7 °C to ensure uniform flowering. Apple trees perform best in deep, well-drained loamy soils with a pH of 5.5–6.5, adequate sunlight, balanced nutrition, proper irrigation, pruning, and effective pollination as shown in Table 1. These conditions collectively support healthy growth, high yield, and superior fruit quality.

Apple Black Root Rot: A soil-borne fungal disease as shown in Figure 1 that causes root decay, leaf yellowing, reduced nutrient uptake, and gradual tree decline, particularly under poorly drained soil conditions.

Apple Fruit Diseases: Fruit diseases such as apple scab and fruit rot produce dark lesions, sunken spots, cracking, and decay on fruit surfaces, leading to reduced market value and storage life.

Apple Leaf Diseases: Leaf diseases appear as spots, powdery growth, chlorosis, or necrosis, reducing photosynthetic efficiency and causing premature defoliation and yield loss.

Table 1. Agricultural Requirements for Apple (*Malus domestica*) Production

Parameter	**Optimal Range / Condition**	**Description**
Climate Type	Temperate	Requires a cool temperate climate with distinct seasons for dormancy, flowering, and fruit development.
Optimal Temperature	18–24 °C	Suitable for vegetative growth and fruit development; high temperatures may reduce fruit quality.
Winter Chilling Requirement	800–1500 h (<7 °C)	Essential for breaking dormancy and uniform bud break; varies by cultivar.
Frost Sensitivity	High during flowering	Spring frost can damage blossoms, leading to poor fruit set and yield reduction.
Annual Rainfall	800–1200 mm	Even distribution is ideal; excessive moisture increases disease incidence.
Irrigation Requirement	Moderate to High	Supplemental irrigation needed during dry periods, especially flowering and fruit enlargement.
Soil Type	Deep, well-drained loamy	Supports healthy root growth and efficient nutrient uptake.
Soil pH	5.5–6.5	Slightly acidic soils enhance nutrient availability.
Soil Drainage	Well-drained	Poor drainage causes waterlogging and root diseases.
Sunlight Requirement	Full sunlight (6–8 h/day)	Required for photosynthesis, fruit color, and sugar accumulation.
Nutrient Requirement	N, P, K, Ca, B, Zn	Balanced nutrition essential; calcium improves fruit firmness and storage life.
Pollination	Cross-pollination	Most cultivars require cross-pollination, often aided by honeybees.
Plant Spacing	3–5 m	Depends on rootstock and training system; improves air circulation.
Pruning Requirement	Annual pruning	Enhances canopy structure, light penetration, and fruit quality.
Major Diseases	Scab, powdery mildew, fire blight	Requires regular monitoring and disease management.
Harvesting Season	Late summer–autumn	Proper timing affects fruit quality, shelf life, and market value.

Other Major Diseases: Diseases including powdery mildew and fire blight affect shoots, flowers, and fruits, resulting in tissue deformation, wilting, and necrosis under favorable environmental conditions.

1.2 Banana Crop and Disease Overview

Fig. 2. Banana disease and Healthy sample of ABCGMP datasets [23].

Banana (*Musa spp.*) is an important tropical fruit crop cultivated for its nutritional value, economic importance, and year-round production. Banana plants grow best in warm and humid climates with an optimal temperature range of 25–30 °C, high relative humidity, and evenly distributed rainfall. The crop performs well in deep, well-drained loamy or alluvial soils with a pH of 6.0–7.5 and requires balanced nutrition, adequate irrigation, wind protection, and proper orchard management to achieve high yield and fruit quality as shown in Table 2.

Banana Cordana Leaf Spot: A fungal leaf disease as shown in Figure 2 characterized by elongated brown to dark lesions with yellow halos, leading to premature leaf drying and reduced photosynthetic efficiency.

Banana Fruit Diseases: Fruit diseases cause surface blemishes, uneven ripening, soft rot, and internal discoloration, significantly reducing market value and post-harvest shelf life.

Banana Leaf Diseases: Leaf diseases such as Sigatoka appear as streaks, necrotic spots, and chlorotic patches, weakening plants and negatively affecting bunch development.

Other Major Diseases: Diseases including Panama disease and bacterial wilt affect roots and vascular tissues, causing plant wilting, reduced vigor, and severe yield losses.

1.3 Citrus Crop and Disease Overview

Citrus (*Citrus spp.*) is a commercially important group of fruit crops, including orange, lemon, lime, mandarin, and grapefruit, valued for their nutritional and economic significance. Citrus trees grow best in tropical and subtropical climates

Table 2. Agricultural Requirements for Banana (*Musa spp.*) Production

Parameter	**Optimal Range / Condition**	**Description**
Climate Type	Tropical to Subtropical	Requires warm and humid climatic conditions for continuous growth.
Optimal Temperature	25–30 °C	Ideal for vegetative growth and fruit development; extreme temperatures reduce yield.
Relative Humidity	>60%	High humidity promotes leaf expansion and bunch formation.
Rainfall Requirement	1000–2500 mm annually	Well-distributed rainfall is essential; supplemental irrigation is required during dry periods.
Irrigation Requirement	High	Frequent irrigation is needed due to shallow root systems and high water demand.
Soil Type	Deep, well-drained loamy or alluvial soil	Fertile soils with good moisture retention support healthy growth.
Soil pH	6.0–7.5	Slightly acidic to neutral soils are optimal for nutrient availability.
Soil Drainage	Well-drained	Poor drainage leads to waterlogging and root rot.
Sunlight Requirement	Full sunlight	Essential for photosynthesis and high fruit yield.
Nutrient Requirement	N, P, K, Mg, Zn, B	Banana is a heavy feeder; potassium is critical for fruit size and quality.
Wind Sensitivity	High	Strong winds can cause lodging; windbreaks are recommended.
Plant Spacing	1.8–2.5 m	Depends on cultivar and planting density; improves air circulation.
Propagation Method	Suckers / Tissue culture	Tissue culture ensures uniform, disease-free planting material.
Major Diseases	Panama disease, Sigatoka leaf spot	Effective disease management is essential for sustained productivity.
Harvesting Period	10–15 months after planting	Harvest duration varies by cultivar and climatic conditions.

Fig. 3. Citrus disease and Healthy sample of ABCGMP datasets [23].

Table 3. Agricultural Requirements for Citrus (*Citrus spp.*) Production

Parameter	**Optimal Range / Condition**	**Description**
Climate Type	Tropical to Subtropical	Suitable for evergreen growth and continuous fruiting cycles.
Optimal Temperature	20–30 °C	Ideal for vegetative growth, flowering, and fruit development.
Frost Sensitivity	High	Frost damages flowers, fruits, and young shoots, reducing yield.
Rainfall Requirement	600–1200 mm annually	Even rainfall is ideal; supplemental irrigation is required during dry periods.
Irrigation Requirement	Moderate	Regular irrigation supports fruit development and prevents moisture stress.
Soil Type	Deep, well-drained loamy or sandy loam soil	Ensures proper root growth, aeration, and nutrient uptake.
Soil pH	5.5–7.5	Slightly acidic to neutral soils favor nutrient availability.
Soil Drainage	Well-drained	Poor drainage increases the risk of root diseases and tree decline.
Sunlight Requirement	Full sunlight	Essential for photosynthesis, flowering, and fruit color development.
Nutrient Requirement	N, P, K, Ca, Mg, Zn, Fe	Balanced fertilization is necessary for healthy growth and fruit quality.
Pruning Requirement	Light annual pruning	Maintains canopy structure, airflow, and light penetration.
Plant Spacing	4–6 m	Depends on species and rootstock; ensures adequate canopy development.
Major Diseases	Citrus canker, greening (HLB), root rot	Disease management is critical for orchard longevity.
Propagation Method	Budding / Grafting	Ensures true-to-type plants and early fruiting.
Harvesting Season	Varies by species	Fruit maturity depends on citrus type and climatic conditions.

with an optimal temperature range of 20–30 °C and adequate sunlight. Deep, well-drained loamy or sandy loam soils with a pH of 5.5–7.5, balanced nutrition, proper irrigation, and effective orchard management are essential for achieving high yield and fruit quality as shown in Table 3. Citrus trees are sensitive to frost and waterlogging, which can severely affect productivity.

Citrus Canker: A bacterial disease as shown in Figure 3 causing raised, corky lesions with yellow halos on leaves, stems, and fruits, leading to defoliation, fruit drop, and reduced market quality.

Citrus Fruit Diseases: Fruit diseases result in rind blemishes, sunken lesions, cracking, soft rot, and discoloration, significantly reducing fruit quality, shelf life, and market value.

Citrus Leaf Diseases: Leaf diseases appear as spots, chlorosis, mottling, and necrotic patches, reducing photosynthesis and weakening tree vigor.

Other Major Diseases: Diseases such as citrus greening (HLB) and root rot affect vascular and root systems, causing leaf yellowing, tree decline, poor fruit development, and severe yield loss.

1.4 Guava Crop and Disease Overview

Fig. 4. Guava disease and Healthy sample of ABCGMP datasets [23].

Guava (*Psidium guajava*) is a widely grown tropical and subtropical fruit crop valued for its high nutritional content and adaptability to diverse agro-climatic conditions. Guava grows best at temperatures between 20–35 °C in deep, well-drained loamy soils with a pH of 5.5–7.0. Adequate soil moisture during flowering and fruit development, balanced nutrition, pruning, and proper orchard management are essential for achieving high yield and good fruit quality as shown in Table 4. Young plants are sensitive to frost, which can negatively affect growth and productivity.

Guava Red Rust: A fungal disease characterized by reddish-brown powdery spots on leaves and young shoots, leading to reduced photosynthesis and weakened plant growth as shown in Figure 4.

Table 4. Agricultural Requirements for Guava (*Psidium guajava*) Production

Parameter	Optimal Range / Condition	Description
Climate Type	Tropical to Subtropical	Adaptable to diverse climatic conditions with moderate temperatures.
Optimal Temperature	20–35 °C	Suitable for vegetative growth, flowering, and fruit development.
Frost Sensitivity	Moderate to High	Young plants are particularly sensitive to frost injury.
Rainfall Requirement	600–1000 mm annually	Moderate rainfall is ideal; excessive rain may affect fruit quality.
Irrigation Requirement	Moderate	Irrigation is essential during flowering and fruit development stages.
Soil Type	Deep, well-drained loamy soil	Supports healthy root growth and nutrient uptake.
Soil pH	5.5–7.0	Slightly acidic to neutral soils favor nutrient availability.
Soil Drainage	Well-drained	Waterlogging leads to root diseases and reduced productivity.
Sunlight Requirement	Full sunlight	Essential for flowering, fruit set, and sugar accumulation.
Nutrient Requirement	N, P, K, Ca, Zn, B	Balanced fertilization improves yield and fruit quality.
Pruning Requirement	Regular pruning	Encourages new shoot growth and improves fruit size and quality.
Plant Spacing	4–6 m	Depends on cultivar and training system.
Major Diseases	Wilt, anthracnose, fruit fly infestation	Integrated pest and disease management is required.
Propagation Method	Air layering / Grafting	Ensures uniform growth and early fruiting.
Harvesting Season	Varies by region	Fruits are harvested at maturity for optimal quality.

Guava Fruit Diseases: Fruit diseases cause surface blemishes, soft rot, cracking, sunken lesions, and discoloration, resulting in reduced market value and post-harvest losses.

Guava Leaf Diseases: Leaf diseases appear as spots, chlorosis, necrosis, and leaf curling, causing premature defoliation and reduced fruit set.

Other Major Diseases: Diseases such as wilt and anthracnose affect roots, stems, leaves, and fruits, leading to plant decline, poor fruit development, and yield reduction.

1.5 Mango Crop and Disease Overview

Table 5. Agricultural Requirements for Mango (*Mangifera indica*) Production

Parameter	**Optimal Range / Condition**	**Description**
Climate Type	Tropical to Subtropical	Requires warm climates with a distinct dry period for flowering.
Optimal Temperature	24–30 °C	Ideal for vegetative growth and fruit development.
Rainfall Requirement	750–2500 mm annually	Dry weather during flowering and moderate rainfall during fruit growth are preferred.
Irrigation Requirement	Moderate	Irrigation is essential during fruit development; excess moisture should be avoided.
Soil Type	Deep, well-drained loamy soil	Supports strong root growth and efficient nutrient uptake.
Soil pH	5.5–7.5	Slightly acidic to neutral soils favor nutrient availability.
Soil Drainage	Well-drained	Poor drainage leads to root diseases and reduced tree vigor.
Sunlight Requirement	Full sunlight	Essential for flowering, fruit set, and fruit quality.
Nutrient Requirement	N, P, K, Ca, Zn, B	Balanced fertilization improves yield and fruit quality.
Frost Sensitivity	Moderate to High	Young trees are sensitive to frost and cold injury.
Pruning Requirement	Periodic pruning	Maintains canopy structure and improves light penetration.
Plant Spacing	8–10 m	Depends on cultivar and training system.
Major Diseases	Powdery mildew, anthracnose	Disease control is critical during flowering and fruiting stages.
Propagation Method	Grafting	Ensures true-to-type plants and early bearing.
Harvesting Season	Summer	Harvest timing varies by cultivar and region.

Mango (*Mangifera indica*) is a major tropical fruit crop valued for its economic importance and nutritional quality. It grows best in tropical and subtropical regions with warm temperatures ranging from 24–30 °C and a distinct dry period for flowering. Mango performs well in deep, well-drained loamy soils

Fig. 5. Mango disease and Healthy sample of ABCGMP datasets [23].

with a pH of 5.5–7.5 and requires balanced nutrition, proper irrigation, pruning, and orchard management to ensure high yield and fruit quality as shown in Table 5. Young trees are sensitive to frost, which can adversely affect growth and productivity.

Mango Bacterial Canker: A bacterial disease as shown in Figure 5 characterized by dark, water-soaked lesions on leaves, stems, and fruits, often leading to cracking, gummosis, and premature fruit drop.

Mango Fruit Diseases: Fruit diseases cause surface blemishes, sunken lesions, soft rot, uneven ripening, and fungal growth, significantly reducing fruit quality, shelf life, and market value.

Mango Leaf Diseases: Leaf diseases such as powdery mildew and anthracnose appear as powdery growth or dark necrotic spots, reducing photosynthesis and affecting flowering and fruit set.

Other Major Diseases: Diseases including dieback and malformation affect shoots and panicles, leading to distorted growth, poor fruit development, and yield reduction.

1.6 Papaya Crop and Disease Overview

Fig. 6. Papaya disease and Healthy sample of ABCGMP datasets [23].

Table 6. Agricultural Requirements for Papaya (*Carica papaya*) Production

Parameter	**Optimal Range / Condition**	**Description**
Climate Type	Tropical to Subtropical	Requires warm and frost-free climatic conditions.
Optimal Temperature	22–32 °C	Ideal for vegetative growth, flowering, and fruit development.
Frost Sensitivity	Very High	Frost causes severe damage and plant mortality.
Rainfall Requirement	1000–2000 mm annually	Well-distributed rainfall is essential; excess moisture should be avoided.
Irrigation Requirement	Moderate to High	Frequent irrigation is required due to shallow root system.
Soil Type	Light, well-drained loamy or sandy loam soil	Promotes healthy root development and prevents waterlogging.
Soil pH	5.5–7.0	Slightly acidic to neutral soils support nutrient availability.
Soil Drainage	Excellent drainage required	Poor drainage leads to root rot and stem diseases.
Sunlight Requirement	Full sunlight	Essential for rapid growth and high fruit yield.
Nutrient Requirement	N, P, K, Ca, Mg, B, Zn	Heavy feeder; balanced fertilization improves yield and fruit quality.
Wind Sensitivity	High	Strong winds can cause stem breakage; windbreaks are recommended.
Plant Spacing	1.8–2.5 m	Depends on cultivar and planting density.
Propagation Method	Seeds	Healthy, high-quality seeds ensure uniform plant growth.
Major Diseases	Papaya ringspot virus, root rot	Disease management is critical for sustained productivity.
Harvesting Period	8–10 months after planting	Early bearing crop with continuous harvesting potential.

Papaya (*Carica papaya*) is a fast-growing tropical fruit crop valued for its nutritional and medicinal importance and year-round production. It grows best in warm tropical and subtropical climates with an optimal temperature range of 22–32 °C. Papaya performs well in light, well-drained loamy or sandy loam soils with a pH of 5.5–7.0 and requires balanced nutrition, proper irrigation, wind protection, and good drainage as shown in Table 6. The crop is highly sensitive to frost, waterlogging, and strong winds, which can severely affect plant growth and fruit yield.

Papaya Ringspot Disease: A viral disease [2] characterized by mosaic patterns, ring-shaped spots on fruits, leaf distortion, and stunted plant growth, leading to severe yield loss as shown in Figure 6.

Papaya Fruit Diseases: Fruit diseases cause surface blemishes, soft rot, sunken lesions, uneven ripening, and premature fruit drop, reducing market quality and shelf life.

Papaya Leaf Diseases: Leaf diseases appear as chlorosis, necrotic lesions, leaf curling, and deformation, reducing photosynthetic efficiency and weakening plant vigor.

Other Major Diseases: Diseases such as root rot and stem rot affect the root and stem systems, causing wilting, plant decline, and eventual crop failure.

1.7 Literature Review

The advancement of deep learning–based plant disease diagnosis has been closely linked to the evolution of publicly available image datasets. Early progress was driven by the introduction of the PlantVillage dataset in 2015, which demonstrated the effectiveness of deep convolutional neural networks for large-scale plant disease classification [3]. Despite its influence, PlantVillage relied primarily on laboratory-controlled images with homogeneous backgrounds, resulting in significant background bias and limited generalization to real-field agricultural conditions. To mitigate these shortcomings, subsequent studies shifted toward field-acquired imagery. Thapa et al. [4] and Singh et al. [5] highlighted the importance of capturing natural disease manifestations under varying illumination, occlusions, and complex backgrounds to enhance model robustness and real-world applicability. More recently, crop-specific repositories such as SAR-MLD for mango [6] and Yolo-Papaya [7] have emerged, integrating both leaf and fruit images to enable dual-organ disease diagnosis.

Between 2024 and 2026, research efforts increasingly concentrated on high-value horticultural crops, including guava, banana, and citrus, leveraging large-scale datasets and advanced deep learning architectures. In guava (*Psidium guajava* L.), studies evolved from basic leaf-based classification toward hybrid diagnostic frameworks that combine deep transfer learning with ensemble machine learning classifiers such as XGBoost, achieving fruit-level disease detection accuracies of up to 98.6% [14, 17]. Banana (*Musa* spp.) research emphasized extensive field-acquired datasets encompassing complex diseases such as Moko, Panama disease, and Black Sigatoka, where lightweight architectures such as MobileNetV2 demonstrated a favorable balance between accuracy (approximately

96%) and computational efficiency, making them suitable for mobile and edge-based deployment [16]. In citrus crops, instance segmentation approaches based on YOLOv8 enabled real-time localization of disease symptoms on fruits and stems, reporting recall rates exceeding 90% [15]. Collectively, these advancements contribute to reducing post-harvest losses and improving resilience against major biotic stresses.

Parallel progress has been reported across several staple and plantation crops. In 2023, Bhandari et al. achieved an accuracy of 99.07% for tomato disease classification using EfficientNetB5, while Tian et al. reported comparable performance with an InceptionV3-based framework [8, 13]. In 2024, Palaniappan et al. employed ResNet50 for cashew disease detection, attaining an accuracy of 97.76%, whereas Daphal et al. reported a lower accuracy of 86.53% for sugarcane disease detection using an AMRCNN architecture, reflecting the increased complexity of field-level sugarcane disease patterns [12, 9]. For rice crops, Joseph et al. introduced a hybrid deep learning model achieving 99.68% accuracy, while Mizan et al. applied EfficientNet-B3 to Bangladeshi rice disease data and obtained an accuracy of 92.00% [10, 11]. These findings indicate that although advanced CNN and hybrid architectures can achieve excellent performance under controlled or crop-specific conditions, their effectiveness varies substantially across crops, datasets, and environmental settings.

Recent studies [21] have demonstrated the effectiveness of transfer learning and hybrid deep learning architectures for crop disease detection. In the context of rice disease classification, pre-trained models such as ResNet50, MobileNetV2, and MobileNet were evaluated, where a hybrid MobileNet–ResNet concatenation model achieved a training accuracy of 99.50% and a test accuracy of 94.00%. This hybrid architecture exhibited superior generalization and robustness compared to individual backbone models, highlighting the benefits of feature-level integration.

For multi-crop and multi-disease classification [20], a feature-level concatenation framework combining EfficientNetB1, DenseNet121, and InceptionV3 was proposed and evaluated on a chilli–rice disease dataset. The model attained 99.86% training accuracy and 92.05% test accuracy by effectively capturing multi-scale and hierarchical representations, thereby outperforming standalone deep networks.

In another study [18], a Transformer-based attention mechanism was integrated with MobileNetV2 and DenseNet121 for disease classification using the CCMT dataset. The incorporation of attention enhanced contextual feature learning, leading to improved robustness and an overall classification accuracy of 99.30%.

Similarly, a hybrid MobileNet–VGG19 architecture for cashew leaf disease detection [19] achieved a test accuracy of 96.23%, along with strong precision and recall, indicating its effectiveness in capturing discriminative disease patterns.

Furthermore, a novel hybrid model termed EBITRNet-B0 [22] was introduced for papaya and sugarcane leaf disease detection by integrating EfficientNet-B0 and BiT-ResNetv2 with a channel-wise attention mechanism. In this architecture, EfficientNet-B0 extracts lightweight, fine-grained features, while BiT-ResNetv2 provides robust and high-level representations, and the attention module emphasizes disease-relevant regions. When evaluated on the PSMCMD dataset, the proposed model achieved 93.40% training accuracy and 87.23% test accuracy, demonstrating improved efficiency and robustness for multi-crop disease classification.

1.8 Research Gap and Motivation for the ABCGMP Dataset

Despite substantial progress, the existing literature reveals a critical gap: the absence of a unified, large-scale, multi-crop fruit-and-leaf disease dataset collected under real-world field conditions. Most publicly available datasets remain crop-specific, geographically constrained, or limited to either leaf or fruit imagery, thereby restricting the development of generalized and transferable disease recognition models. Furthermore, current benchmarks insufficiently address inter-crop visual similarity, cross-disease confusion, and practical commercial grading requirements. To bridge this gap, this study introduces the **ABCGMP Fruit and Leaf Disease Dataset**, comprising six economically important crops—apple, banana, citrus, guava, mango, and papaya. The proposed dataset is designed to support multi-crop and multi-organ disease diagnosis under diverse environmental conditions, enabling robust benchmarking of convolutional neural networks, transformer-based models, and hybrid attention-driven architectures for real-world agricultural deployment.

Value of the Data

- Facilitates multi-crop and multi-disease plant health assessment.
- Compatible with both lightweight and deep architectures, such as MobileNet and MobileNetV2.
- Supports conventional CNN as well as hybrid CNN-based models.
- Captures real-world agricultural variability, enabling robust AI model evaluation.
- Well-suited for explainable AI (XAI) and uncertainty estimation analyses.

2 ABCGMP Fruit and Leaf Disease Dataset Description

The Apple–Banana–Citrus–Guava–Mango–Papaya (ABCGMP) dataset is a comprehensive multi-crop, multi-disease image repository containing both fruit and leaf samples from six economically important horticultural crops: Apple, Banana, Citrus, Guava, Mango, and Papaya. The dataset includes healthy and diseased samples exhibiting diverse symptoms such as leaf spots, blight, anthracnose, bacterial infections, chlorosis, necrosis, fruit rot, scab-like lesions, and viral manifestations. As shown in Figure 5, mango fruit images display visible necrotic lesions and clustered spots typical of fungal infections under natural field conditions,

Table 7. Crop wise class and image distribution of ABCGMP datasets [23].

S.No	Crops	Class	No. of images	Crop	Class	No. of images
1.	Apple	Healthy Leaf	2008	Guava	Red Rust	90
2.	Apple	Black Rot	1987	Guava	Healthy Leaf	150
3.	Apple	Healthy Fruit	301	Guava	Disease Fruit	160
4.	Apple	Disease Fruit	286	Guava	Healthy Fruit	50
5.	Banana	Cordana	739	Mango	Bacterial Cancker	404
7	Banana	Healthy Leaf	535	Mango	Healthy Leaf	308
8	Banana	Healthy Fruit	133	Mango	Disease Fruit	213
9	Banana	Disease Fruit	74	Mango	Healthy Fruit	348
10	Citrus	Cancker Leaf	210	Papaya	Ring Spot	717
11	Citrus	Healthy Leaf	192	Papaya	Healthy Leaf	298
12	Citrus	Disease Fruit	158	Papaya	Disease Fruit	127
13	Citrus	Healthy Fruit	190	Papaya	Healthy Fruit	143

with similar real-world symptom diversity preserved across all crops as shown in Table 7.

To enhance robustness and real-world applicability, images were collected from heterogeneous sources. Approximately 50% were acquired directly from agricultural fields under uncontrolled conditions, capturing variations in illumination, background, and disease severity. Around 16% were contributed by farmers using smartphone cameras across different geographic locations. Additionally, 20% were obtained from publicly available benchmark plant disease datasets, while the remaining 14% were sourced from verified online agricultural repositories and carefully curated to ensure accurate labeling and visual quality as shown in Table 8.

Table 8. Source-wise distribution of images in the ABCGMP datasets [23].

S.No	**Data Source**	**Description**	**(%)**
1	Field collection	Images captured directly from farms under natural field conditions	50
2	Farmer contributions	Images provided by farmers using mobile devices	16
3	Public datasets	Images collected from established benchmark datasets	20
4	Verified internet sources	Carefully screened agricultural images from online sources	14
		Total	**100**

A substantial proportion of the field-acquired images were collected from agriculturally prominent regions of Rajasthan and Jammu, India, covering multiple districts in Jammu as well as the Jaipur, Kota, and Bundi districts. The inclusion of diverse agro-climatic zones and disease prevalence conditions significantly strengthens the dataset's representativeness and generalization capability.

Table 9. Quantitative image acquisition and visual properties of the ABCGMP datasets [23].

S.No	Image Property	Unit / Metric	Observed Range / Value
1	Image Resolution	Pixels (H $\times$ W)	256×256 to 512×512
2	Image Channels	Count	3 (RGB)
3	Image DPI	Dots per Inch	72 – 300 DPI
4	Color Depth	Bits per Channel	8-bit (24-bit RGB)
5	Mean Brightness	Intensity (0–255)	88.2 – 182.5
6	Brightness Std. Dev.	$\sigma_{\text{Intensity}}$	25.4 – 68.3
7	Image Contrast	RMS Contrast	0.19 – 0.67
8	Color Variance	RGB Variance	390 – 2100
9	Shannon Entropy	Bits	5.0 – 7.9
10	Blur Level	Var. of Laplacian	75 – 1320
11	Noise Level	Estimated SNR	17 – 40 dB
12	Occlusion Ratio	Area Coverage	4% – 48%
13	Background Coverage	Image Area	28% – 72%
14	Object Scale (Fruit/Leaf)	Area Ratio	10% – 90%
15	Illumination Condition	Category Count	4 (Bright, Shadowed, Mixed, Low-light)
16	Capture Distance	Approx. Range	20 – 150 cm
17	Compression Format	File Type	JPEG / PNG
18	JPEG Quality Factor	Scale (0–100)	60 – 95
19	Acquisition Device	Category	Smartphone / Digital Camera

All images were captured under varying illumination conditions (bright sunlight, partial shadow, mixed lighting, and low-light), viewpoints, and background settings such as soil, sky, surrounding vegetation, and human interference as shown in Table 9. The deliberate inclusion of such real-world variability supports the development of robust, illumination-invariant deep learning models for multi-crop fruit and leaf disease classification.

Overall, the six-crop dataset provides a challenging benchmark for evaluating deep learning–based classification, detection, and attention-based models for joint fruit and leaf disease recognition in heterogeneous agricultural environments.

3 Experimental Design, Materials and Methods

3.1 Dataset Formulation

Let the dataset be defined as:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}, \tag{1}$$

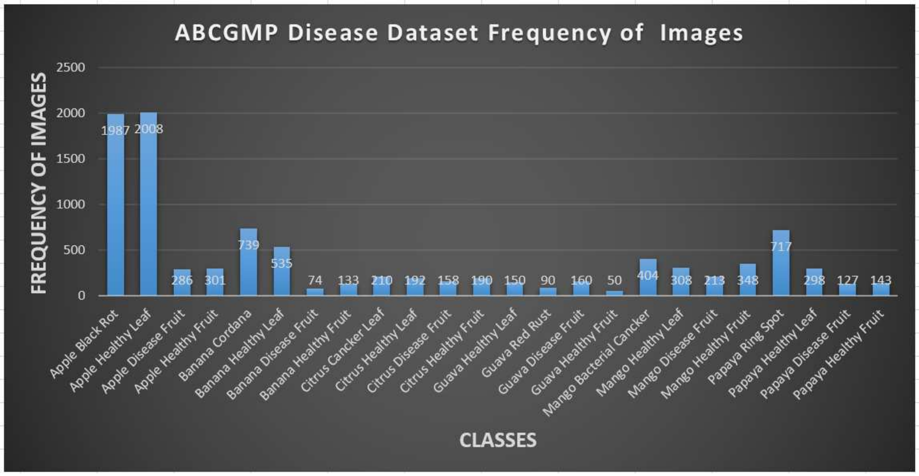

Fig. 7. ABCGMP datasets [23] Frequency of Image.

where $x_i \in \mathbb{R}^{H\times W\times 3}$ is an RGB image and $y_i \in \{1, 2, \ldots, C\}$ denotes the disease class label.

Algorithm 1: CNN-Based Disease Classification on ABCGMP datasets [23].

Input: ABCGMP fruit and leaf disease image dataset
Output: Trained CNN model and disease predictions
Load the ABCGMP dataset and resize all images to 224×224. Split the dataset into training and validation sets using an 80:20 ratio. Normalize pixel values to $[0, 1]$ and apply dataset prefetching for efficient training.
Initialize a pretrained MobileNetV2 model with ImageNet weights and remove the top layer. Construct the CNN by adding Global Average Pooling, Dropout (0.2), and a softmax classification layer. Compile the model using the Adam optimizer and sparse categorical cross-entropy loss.
Train the model for 50 epochs with validation monitoring. Evaluate performance using accuracy, loss curves, a confusion matrix, and a classification report. Save the trained model for future deployment.

3.2 CNN-Based Learning Framework

The CNN-based deep learning framework is implemented using the TensorFlow and Keras libraries for automatic plant disease classification on the proposed **ABCGMP fruit and leaf disease dataset**. The dataset consists of images from multiple crops, and all input images are uniformly resized to 224×224

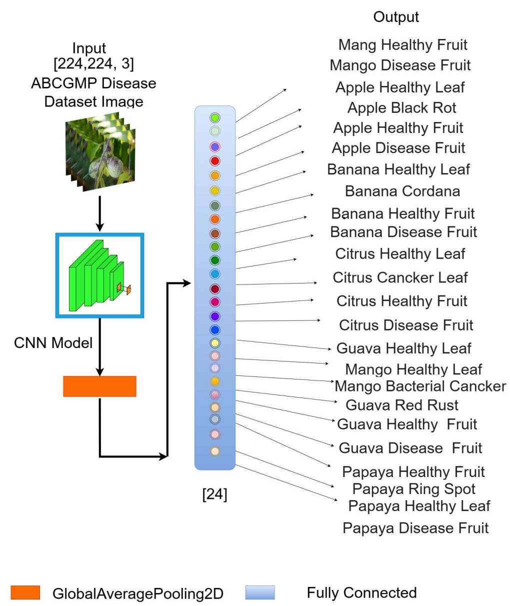

Fig. 8. CNN Model Diagram on ABCGMP datasets [23]..

pixels to ensure compatibility with pretrained convolutional neural network architectures as illustrated in Figure 8. The dataset is partitioned into training and validation sets using an 80:20 split to enable unbiased performance evaluation.

Prior to training, pixel intensity values are normalized to the range $[0, 1]$ using a rescaling operation. To enhance training efficiency, dataset prefetching with $AUTOTUNE$ is employed, allowing optimized data loading and improved GPU utilization. The core of the framework is a CNN model based on a pretrained backbone initialized with ImageNet weights, which serves as a robust feature extractor for capturing discriminative disease-related patterns from fruit and leaf images. The pretrained convolutional layers are kept frozen to retain learned low-level and high-level visual representations.

To reduce the dimensionality of the extracted feature maps, a Global Average Pooling layer is applied, followed by a dropout layer with a rate of 0.2 to alleviate overfitting. The final classification layer employs a softmax activation function to predict multiple disease categories present in the ABCGMP dataset. The model is trained using the Adam optimizer with a learning rate of 10^{-3} and sparse categorical cross-entropy loss over 50 epochs. Model performance is evaluated using accuracy and loss curves, along with a confusion matrix and detailed classification report to assess class-wise predictive performance, as summarized in Algorithm 1.

3.3 Dataset Preparation and Preprocessing

Let the image dataset be defined as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}, \tag{2}$$

where $x_i \in \mathbb{R}^{224\times224\times3}$ denotes an RGB fruit or leaf image and $y_i \in \{1, 2, \ldots, C\}$ represents the corresponding class label, with C being the total number of disease classes.

The dataset is partitioned into training and validation subsets as

$$\mathcal{D}_{\text{train}} = 0.8\,\mathcal{D}, \quad \mathcal{D}_{\text{val}} = 0.2\,\mathcal{D}. \tag{3}$$

Each image is normalized using min–max scaling:

$$x_i^{(\text{norm})} = \frac{x_i}{255}. \tag{4}$$

3.4 Feature Extraction Using CNN Backbone

A pretrained convolutional neural network without its classification head is utilized as a feature extractor. Let $f_\theta(\cdot)$ denote the CNN backbone with frozen parameters θ. The extracted feature maps are computed as

$$\mathbf{F}_i = f_\theta\left(x_i^{(\text{norm})}\right), \tag{5}$$

where $\mathbf{F}_i \in \mathbb{R}^{H\times W\times K}$ represents the high-level spatial feature maps.

3.5 Global Average Pooling

Global Average Pooling (GAP) is applied to aggregate spatial information and reduce feature dimensionality:

$$z_{ik} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} F_{i,h,w,k}, \tag{6}$$

yielding a compact feature vector

$$\mathbf{z}_i = [z_{i1}, z_{i2}, \ldots, z_{iK}] \in \mathbb{R}^K. \tag{7}$$

3.6 Dropout Regularization

To alleviate overfitting, dropout regularization is applied with a probability $p = 0.2$:

$$\tilde{\mathbf{z}}_i = \mathbf{z}_i \odot \mathbf{m}, \quad m_j \sim \text{Bernoulli}(1-p), \tag{8}$$

where $\odot$ denotes element-wise multiplication.

3.7 Classification Layer

The final fully connected layer computes the class logits as

$$\mathbf{o}_i = W\tilde{\mathbf{z}}_i + \mathbf{b}, \tag{9}$$

where $W \in \mathbb{R}^{C \times K}$ is the weight matrix and $\mathbf{b} \in \mathbb{R}^C$ is the bias vector.

The softmax activation function converts logits into class probabilities:

$$\hat{y}_{ic} = \frac{e^{o_{ic}}}{\sum_{j=1}^{C} e^{o_{ij}}}, \quad c = 1, 2, \ldots, C. \tag{10}$$

3.8 Loss Function

Model optimization is guided by the sparse categorical cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \log \left(\hat{y}_{i,y_i} \right). \tag{11}$$

3.9 Optimization Strategy

The network parameters are optimized using the Adam optimizer:

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{12}$$

where $\alpha = 10^{-3}$ denotes the learning rate, and $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates, respectively.

3.10 Prediction and Performance Evaluation

The predicted class label for each input sample is obtained as

$$\hat{y}_i = \arg\max_c \hat{y}_{ic}. \tag{13}$$

Overall model performance is measured using classification accuracy:

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \nVdash(\hat{y}_i = y_i). \tag{14}$$

The confusion matrix is defined as

$$\text{CM}_{p,q} = \sum_{i=1}^{N} \nVdash(y_i = p \wedge \hat{y}_i = q), \tag{15}$$

where p and q denote the true and predicted classes, respectively.

Class-wise precision, recall, and F_1-score are computed as

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \tag{16}$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c}, \tag{17}$$

$$F1_c = \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}. \tag{18}$$

3.11 MobileNetV2 Model

Table 10. MobileNetV2 Model precision, Recall, and F_1-score comparison on ABCGMP datasets [23].

Diseases	Precision (%)	Recall (%)	F_1-Score (%)
Apple Black rot	100	97	98
Apple healthy	99	91	95
Banana cordana	51	80	63
Citrus CanckerLeaf	97	81	88
Citrus Fruit disease	72	86	78
Citrus Healthy Fruit	58	42	49
Citrus Healthy Leaf	47	42	44
Disease Apple Fruit	15	04	07
Disease Banana Fruit	21	40	28
Disease Mango Fruit	80	37	51
Guava Disease Fruit	31	50	39
Guava Healthly Fruit	75	33	46
Guava Healthly Leaf	20	100	34
Guava Red Rust	50	33	40
Healthy Apple Fruit	22	12	16
Healthy Banana Fruit	47	80	59
Healthy Banana Leaf	86	100	93
Healthy Mango Fruit	83	55	66
Mang Healthy Leaf	88	100	94
Mango Bacterial Canker	99	95	97
Papaya Fruit Disease	87	41	55
Papaya Fruit Healthy	85	44	58
Papaya Healthy leaf	86	95	90
Papaya RingSpot	100	58	73
Accuracy			77
Macro Avg	67	62	61
Weighted Avg	82	77	78

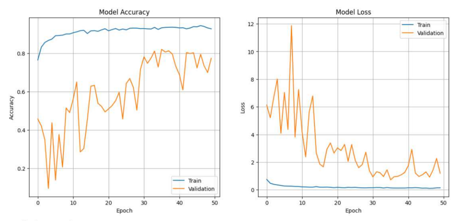

Fig. 9. MobileNetV2 Model Accuracy and Loss graph on ABCGMP datasets [23].

Table 10 summarizes the class-wise performance of the MobileNetV2 model on the ABCGMP dataset. The model achieves high precision, recall, and F_1-scores for several visually distinct classes, including *Apple Black Rot*, *Mango Bacterial Canker*, *Healthy Banana Leaf*, *Mango Healthy Leaf*, and *Papaya Healthy Leaf*, highlighting its ability to learn discriminative features for well-separated disease patterns. Lower performance is observed for fruit-based and visually ambiguous classes such as *Disease Apple Fruit*, *Healthy Apple Fruit*, and *Citrus Healthy Leaf*, mainly due to class imbalance and visual similarity. Overall, the model achieves an accuracy of 77% with a weighted F_1-score of 78%, while the macro-average F_1-score of 61% reflects class-wise variability. Figure 9 shows effective convergence with moderate validation fluctuations, indicating reasonable generalization despite mild overfitting.

3.12 VGG19 Model

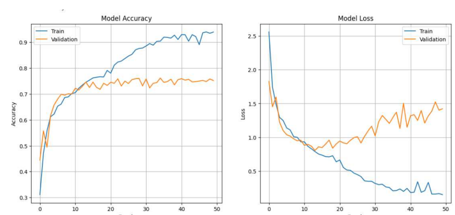

Fig. 10. VGG19 Model Accuracy and Loss graph on ABCGMP datasets [23].

Table 11 summarizes the class-wise performance of the VGG19 model on the ABCGMP dataset. The model achieves very high precision, recall, and F_1-scores for several leaf-based classes, including *Apple Black Rot*, *Apple Healthy*, *Healthy Banana Leaf*, *Mango Healthy Leaf*, and *Mango Bacterial Canker*, demonstrating the effectiveness of deep hierarchical feature extraction for visually consistent disease patterns. Strong performance is also observed for *Papaya RingSpot* and

Table 11. VGG19 Model precision, Recall, and F_1-score comparison on ABCGMP datasets [23].

Diseases	Precision (%)	Recall (%)	F_1-Score (%)
Apple Black rot	99	100	99
Apple healthy	100	99	99
Banana cordana	42	66	51
Citrus CanckerLeaf	76	86	81
Citrus Fruit disease	33	61	43
Citrus Healthy Fruit	33	09	14
Citrus Healthy Leaf	38	42	40
Disease Apple Fruit	10	13	11
Disease Banana Fruit	20	10	13
Disease Mango Fruit	44	16	24
Guava Disease Fruit	29	16	20
Guava Healthly Fruit	20	33	25
Guava Healthly Leaf	66	89	76
Guava Red Rust	50	33	40
Healthy Apple Fruit	05	03	03
Healthy Banana Fruit	23	15	18
Healthy Banana Leaf	98	100	99
Healthy Mango Fruit	35	25	29
Mang Healthy Leaf	99	93	96
Mango Bacterial Canker	99	100	99
Papaya Fruit Disease	35	22	27
Papaya Fruit Healthy	42	20	27
Papaya Healthy leaf	73	85	78
Papaya RingSpot	94	82	88
Accuracy			75
Macro Avg	53	51	50
Weighted Avg	75	75	74

Papaya Healthy Leaf. In contrast, lower performance is noted for fruit-based and visually ambiguous classes such as *Disease Apple Fruit*, *Healthy Apple Fruit*, *Citrus Healthy Fruit*, and *Disease Banana Fruit*, mainly due to class imbalance and high inter-class similarity. Overall, VGG19 achieves an accuracy of 75% with a weighted F_1-score of 74%, while the macro-average F_1-score of 50% highlights class-wise variability. Figure 10 indicates overfitting, as training performance improves steadily while validation accuracy saturates, suggesting limited generalization despite strong results for well-defined classes.

3.13 DenseNet121 Model

Table 12. DenseNet121 Model precision, Recall, and F_1-score comparison on ABCGMP datasets [23].

Diseases	Precision (%)	Recall (%)	F_1-Score (%)
Apple Black rot	100	100	100
Apple healthy	100	98	99
Banana cordana	51	67	58
Citrus CanckerLeaf	64	95	76
Citrus Fruit disease	96	61	75
Citrus Healthy Fruit	51	58	54
Citrus Healthy Leaf	70	70	70
Disease Apple Fruit	11	13	12
Disease Banana Fruit	57	40	47
Disease Mango Fruit	40	88	55
Guava Disease Fruit	67	31	43
Guava Healthly Fruit	80	44	57
Guava Healthly Leaf	82	96	89
Guava Red Rust	100	27	42
Healthy Apple Fruit	04	03	03
Healthy Banana Fruit	73	40	52
Healthy Banana Leaf	99	100	100
Healthy Mango Fruit	59	73	65
Mang Healthy Leaf	96	100	98
Mango Bacterial Canker	100	62	77
Papaya Fruit Disease	50	09	16
Papaya Fruit Healthy	76	52	62
Papaya Healthy leaf	92	92	92
Papaya RingSpot	99	96	97
Accuracy			81
Macro Avg	71	63	64
Weighted Avg	82	81	80

Table 12 summarizes the class-wise performance of the DenseNet121 model on the ABCGMP dataset. The model achieves near-perfect precision, recall, and

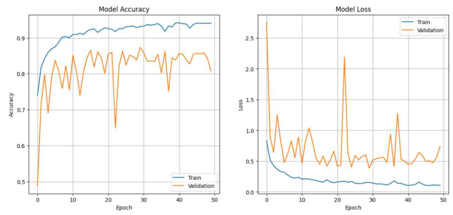

Fig. 11. DenseNet121 Model Accuracy and Loss graph on ABCGMP datasets [23].

F_1-scores for several leaf-based classes, including *Apple Black Rot*, *Apple Healthy*, *Healthy Banana Leaf*, *Mango Healthy Leaf*, and *Papaya RingSpot*, demonstrating the effectiveness of dense connectivity in preserving fine-grained spatial and texture features. Strong results are also observed for *Guava Healthy Leaf*, *Papaya Healthy Leaf*, and *Citrus Canker Leaf*. Lower performance is noted for visually challenging and minority fruit-based classes such as *Disease Apple Fruit*, *Healthy Apple Fruit*, and *Papaya Fruit Disease*, mainly due to class imbalance and subtle symptom variations. Overall, DenseNet121 achieves an accuracy of 81% with a weighted F_1-score of 80%, while the macro-average F_1-score of 64% reflects moderate class-wise variability. Figure 11 indicates stable convergence and good generalization with limited overfitting.

3.14 InceptionV3 Model

TTable 13 summarizes the class-wise performance of the InceptionV3 model on the ABCGMP dataset. The model achieves very high precision, recall, and F_1-scores for several leaf-based classes, including *Apple Black Rot*, *Apple Healthy*, *Healthy Banana Leaf*, *Mango Healthy Leaf*, and *Mango Bacterial Canker*, demonstrating its effectiveness in capturing multi-scale spatial and texture features. Strong performance is also observed for *Citrus Canker Leaf*, *Papaya RingSpot*, and *Guava Healthy Leaf*. In contrast, lower performance is noted for visually ambiguous and minority fruit-based classes such as *Disease Apple Fruit*, *Disease Banana Fruit*, *Citrus Healthy Fruit*, and *Guava Disease Fruit*, mainly due to class imbalance and subtle symptom differences. Overall, the model attains an accuracy of 81% with a weighted average F_1-score of 81%, while the macro-average F_1-score of 65% reflects variability across underrepresented classes. Figure 12 shows stable training convergence and consistent validation performance, indicating good generalization with limited overfitting.

3.15 ResNet50 Model

Table 14 summarizes the class-wise performance of the ResNet50 model on the ABCGMP dataset. The model achieves near-perfect precision, recall, and F_1-scores for several leaf-based classes, including *Apple Black Rot*, *Apple Healthy*,

Table 13. InceptionV3 Model precision, Recall, and F_1-score comparison on ABCGMP datasets [23].

Diseases	Precision (%)	Recall (%)	F_1-Score (%)
Apple Black rot	99	99	99
Apple healthy	100	100	100
Banana cordana	68	53	60
Citrus CanckerLeaf	93	76	84
Citrus Fruit disease	95	58	72
Citrus Healthy Fruit	39	39	39
Citrus Healthy Leaf	29	77	42
Disease Apple Fruit	19	18	18
Disease Banana Fruit	17	30	21
Disease Mango Fruit	72	49	58
Guava Disease Fruit	70	22	33
Guava Healthly Fruit	100	44	62
Guava Healthly Leaf	77	82	79
Guava Red Rust	86	40	55
Healthy Apple Fruit	22	27	25
Healthy Banana Fruit	62	80	70
Healthy Banana Leaf	99	100	100
Healthy Mango Fruit	75	71	73
Mang Healthy Leaf	97	100	99
Mango Bacterial Canker	100	95	97
Papaya Fruit Disease	69	62	66
Papaya Fruit Healthy	73	32	44
Papaya Healthy leaf	56	97	71
Papaya RingSpot	98	85	91
Accuracy			81
Macro Avg	71	64	65
Weighted Avg	84	81	81

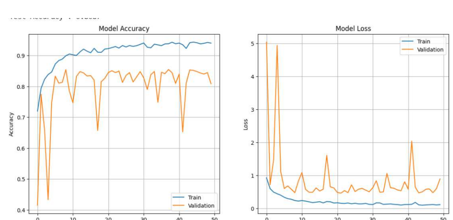

Fig. 12. InceptionV3 Model Accuracy and Loss graph on ABCGMP datasets [23].

Table 14. ResNet50 Model precision, Recall, and F_1-score comparison on ABCGMP datasets [23].

Diseases	Precision (%)	Recall (%)	F_1-Score (%)
Apple Black rot	100	100	100
Apple healthy	100	100	100
Banana cordana	54	62	57
Citrus CanckerLeaf	92	92	92
Citrus Fruit disease	82	75	78
Citrus Healthy Fruit	47	52	49
Citrus Healthy Leaf	60	63	61
Disease Apple Fruit	17	09	12
Disease Banana Fruit	29	40	33
Disease Mango Fruit	72	49	58
Guava Disease Fruit	41	38	39
Guava Healthly Fruit	62	56	59
Guava Healthly Leaf	83	89	86
Guava Red Rust	82	60	69
Healthy Apple Fruit	02	01	01
Healthy Banana Fruit	57	85	68
Healthy Banana Leaf	99	100	100
Healthy Mango Fruit	76	70	73
Mang Healthy Leaf	100	100	100
Mango Bacterial Canker	98	100	99
Papaya Fruit Disease	59	62	61
Papaya Fruit Healthy	52	68	59
Papaya Healthy leaf	89	92	90
Papaya RingSpot	97	97	97
Accuracy			83
Macro Avg	69	69	68
Weighted Avg	83	83	83

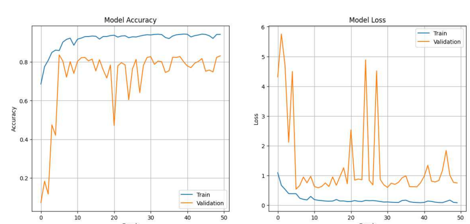

Fig. 13. ResNet50 Model Accuracy and Loss graph on ABCGMP datasets [23].

Healthy Banana Leaf, *Mango Healthy Leaf*, and *Mango Bacterial Canker*, demonstrating the effectiveness of residual learning in capturing deep discriminative features. Strong results are also observed for *Citrus Canker Leaf*, *Papaya RingSpot*, and *Guava Healthy Leaf*. Lower performance is noted for visually ambiguous and minority fruit-based classes such as *Disease Apple Fruit* and *Disease Banana Fruit*, mainly due to class imbalance and subtle symptom variations. Overall, the model achieves an accuracy of 83% with weighted precision, recall, and F_1-score of 83%, while a macro-average F_1-score of 68% indicates moderate class-wise variability. Figure 13 shows stable convergence with minor validation fluctuations, suggesting good generalization with limited overfitting.

Table 15. Model Accuracy & #Parameters on ABCGMP datasets [23]..

Model	Accuracy (%) (%)	Test Accuracy (%)	#Training Parameters (Million)	#Non Taining Parameters (Million)	Epoch Time (Second)	Model Weight
MobileNetV2	92.16	77.37	2.25	0.03	20	1280
DenseNet121	93.91	80.77	6.97	0.08	43	1024
ResNet50	93.82	83.11	23.58	0.05	41	2048
InceptionV3	93.95	80.87	21.81	0.03	37	2048
VGG19	93.33	75.19	20.03	0.00	73	512

Table 15 presents a comparative analysis of five widely used deep learning architectures in terms of accuracy, model complexity, and computational cost. MobileNetV2 achieves a training accuracy of 92.16% and a test accuracy of 77.37% while remaining lightweight, with only 2.25 million trainable parameters and a short training time of 20 seconds, making it suitable for resource-constrained environments. DenseNet121 improves performance with a training accuracy of 93.91% and a test accuracy of 80.77%, at the expense of increased model size (6.97 million parameters) and training time (43 seconds). ResNet50 demonstrates strong generalization capability by achieving the highest test accuracy of 83.11%, though it requires substantially more parameters (23.58 million). InceptionV3 provides a balanced trade-off between accuracy (93.95% training and 80.87% test accuracy) and computational complexity, whereas VGG19, despite having the longest training time (73 seconds), shows comparatively lower test accuracy (75.19%). Overall, the results highlight a clear trade-off between accuracy and efficiency, where lightweight models favor faster training and reduced complexity, while deeper architectures tend to achieve better generalization.

3.16 Features Map Analysis

The MobileNetV2 feature map visualizations reveal a clear hierarchical learning behavior for both Mango and Papaya images as shown in Figure 14and Figure 15. In the early layers, the feature maps strongly emphasize low-level visual

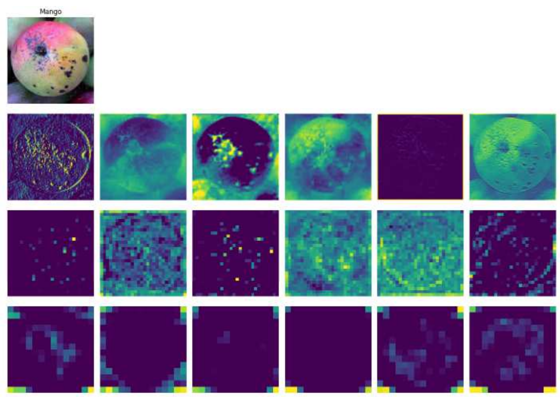

Fig. 14. MobileNrtV2 Model Features analysis on Disease Mango Fruit sample of ABCGMP datasets [23].

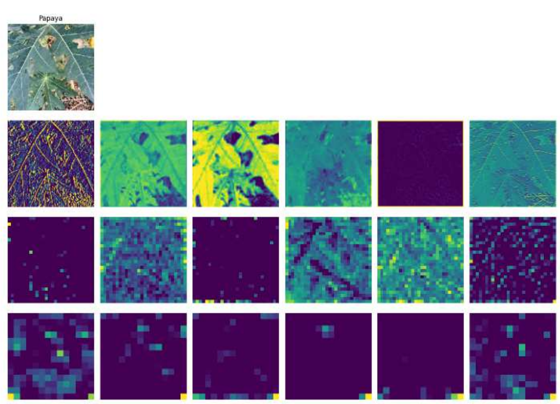

Fig. 15. MobileNrtV2 Model Features analysis on Healthy Papaya Leaf sample of ABCGMP datasets [23].

cues such as edges, contours, color gradients, and surface textures, which are evident from the dense and high-frequency activations outlining fruit boundaries, leaf veins, and lesion edges. As the network progresses to intermediate layers, the activations become more localized and structured, capturing disease-relevant patterns such as dark spots on mango skin and irregular lesion regions on papaya leaves, while suppressing background noise. In the deeper layers, the feature maps appear more sparse and abstract, indicating that MobileNetV2 is focusing on high-level semantic representations that correspond to discriminative disease characteristics rather than raw visual details. This gradual transition from fine-grained texture extraction to compact, class-specific feature encoding demonstrates the efficiency of MobileNetV2's inverted residual and depthwise separable convolution design in capturing salient disease features with minimal redundancy, making it well-suited for lightweight and accurate plant disease classification.

3.17 Heatmap Uncertainty

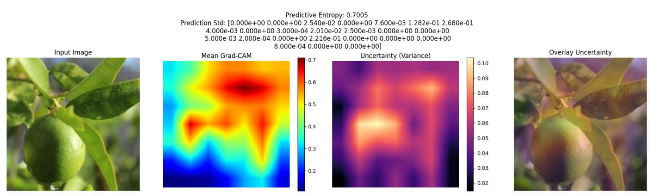

Fig. 16. MobileNrtV2 Model Heat Map Uncertainity analysis on Disease Mango Fruit sample of ABCGMP datasets [23].

The MobileNetV2 heatmap uncertainty analysis illustrates how the model balances confident feature localization with regions of predictive ambiguity as shown in Figure 16. The mean Grad-CAM highlights the most discriminative areas of the image, primarily focusing on the fruit body and adjacent leaf regions, indicating that MobileNetV2 relies on these anatomical structures for its final decision. The predictive entropy value reflects a moderate level of uncertainty, suggesting that while the model is reasonably confident, some class overlap or visual similarity may still exist. The uncertainty (variance) map further reveals localized high-variance regions around the fruit surface and leaf boundaries, where variations in texture, illumination, or disease-like patterns can influence the prediction. When overlaid on the input image, these uncertain regions become visually interpretable, showing that ambiguity is concentrated in transition areas rather than the background. Overall, this analysis demonstrates that MobileNetV2 not only attends to biologically meaningful regions but also provides reliable uncertainty cues, which are crucial for trustworthy plant disease diagnosis and decision support in real-world agricultural applications.

3.18 Quantitative Analysis of Correctly Classified and Misclassified Samples Using MobileNet

Fig. 17. Quantitative Analysis of Correctly Classified and Misclassified ABCGMP datasets [23] samples on MobileNetV2 Model.

The MobileNetV2 classification and misclassification analysis highlights both the strengths and limitations of the model when applied to multi-crop and multi-disease recognition as shown in Figure 17. Correctly classified samples demonstrate the model's ability to capture clear visual cues such as uniform color, healthy texture, and well-defined leaf or fruit structures, resulting in accurate predictions for healthy mango fruits, banana leaves, and apple black rot cases. However, the misclassified samples reveal challenges arising from strong inter-class visual similarity, especially between healthy and diseased categories across different crops. In several cases, overlapping color patterns, comparable vein structures, and subtle texture variations lead the model to confuse apple leaves with papaya disease classes or mango fruits with papaya-related categories. Variations in illumination, background clutter, and scale further contribute to these errors. This analysis indicates that while MobileNetV2 effectively learns discriminative features for most samples, misclassifications are primarily driven by ambiguous visual patterns and cross-crop similarities, suggesting that enhanced data diversity, fine-grained feature learning, or attention-based mechanisms could further improve classification robustness.

4 Conclusion

This paper presented a comprehensive multi-crop fruit and leaf disease image dataset covering six economically important crops, namely Apple, Banana, Citrus, Guava, Mango, and Papaya. The dataset includes both healthy and diseased samples collected under real-field and semi-controlled environments, capturing a wide range of disease symptoms, severity levels, growth stages, and background complexities. By incorporating fruit and leaf images across multiple crops, the dataset addresses the limitations of single-crop datasets and supports the development of robust and generalized plant disease analysis systems.

The diversity and variability inherent in the dataset make it particularly suitable for training and evaluating state-of-the-art deep learning models. Modern convolutional neural networks, transformer-based architectures, and hybrid attention-enhanced models can effectively leverage this dataset to learn discriminative visual features related to texture, color variation, lesion morphology, and spatial disease patterns. The dataset enables fair benchmarking of advanced models for multi-class disease classification, cross-crop generalization, and real-world deployment scenarios. Overall, this dataset serves as a valuable resource for the research community, facilitating advancements in automated plant disease detection and contributing toward intelligent, scalable, and sustainable agricultural decision-support systems.

Acknowledgements

Rajsthan Government Agriculture Department and Kota Agriculture University Kota

Declarations

- Funding: No funding
- Conflict of interest/Competing interests: There is no conflict of interest.
- Ethics approval and consent to participate: Not applicable.
- Consent for publication: Yes
- Data availability: Yes
  Data is available on Mendeley:[1]
  Enquiries about data availability should be directed to the authors.
- Materials availability: Yes
- Code availability: Code is available on GitHub [2]
  Enquiries about code availability should be directed to the authors.
- Author contribution: Joint contribution
- Corresponding authors: Anand Kumar Jain
  Corresponding authors Email Id: 2023rcp9026@mnit.ac.in

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1] https://data.mendeley.com/drafts/fhbvmpcyy2
[2] https://github.com/anand1982/ABCGMP-Fruit-and-Leaf-Disease-Dataset

Rights and permissions

References

1. Research Analytics Group, "Evolution of Vision Transformers in Plant Pathology: A Meta-Analysis of Dataset Robustness," *PMC Agricultural AI Series*, vol. 8, no. 2, 2026.
2. Anand, A. K. Jain, N. Nain, and A. Jain, "PSMCMD," *Mendeley Data*, V6, 2026. https://data.mendeley.com/drafts/5555cvmmgb
3. D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.
4. R. Thapa, K. Zhang, N. Snavely, S. Belongie, and A. Khan, "The Plant Pathology 2020 challenge dataset to classify foliar diseases of apples," *arXiv preprint arXiv:2004.11958*, 2020.
5. U. P. Singh, S. S. Jain, R. K. Jain, and R. K. Puri, "Deep learning-based characterization of anthracnose disease in mango leaves," *IEEE Access*, vol. 8, pp. 116254–116263, 2020.
6. S. Ahmed et al., "SAR-MLD: A high-resolution multi-modal dataset for Mango leaf disease detection in real orchard environments," *Journal of Agricultural Informatics*, vol. 16, no. 1, pp. 22–38, 2025.
7. M. R. K. Tusar et al., "Yolo-Papaya: Real-time detection and quality grading of Papaya fruit diseases," *IEEE Access*, vol. 14, pp. 450–468, 2026.
8. Bhandari, M., Shahi, T. B., Neupane, A., and Walsh, K. B. (2023). Botanicx-ai: Identification of tomato leaf diseases using an explanation-driven deep-learning model. *Journal of Imaging*, 9(2), 53.
9. Daphal, S. D., and Koli, S. M. (2024). Enhanced deep learning technique for sugarcane leaf disease classification and mobile application integration. *Heliyon*, 10(8).
10. Joseph, D. S., and Pawar, P. M. (2024). Mobile-Xcep hybrid model for plant disease diagnosis. *Multimedia Tools and Applications*, 1–44.
11. Mizan, M. A. I., Ahmed, M., and Ali, M. H. (2024). Bangladeshi Crop Disease Detection Using Convolutional Neural Network. *In 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 1304–1309. IEEE.
12. Palaniappan, S., and Pazhamalai, K. (2024). ELDA: enhanced linear discriminant analysis for cashew crop disease detection using precision agriculture. *Journal of Electronic Imaging*, 33(2), 023030.
13. Tian, K., Zeng, J., Song, T., Li, Z., Evans, A., and Li, J. (2023). Tomato leaf diseases recognition based on deep convolutional neural networks. *Journal of Agricultural Engineering*, 54(1).
14. Almadhor, A., Rauf, H., Lali, M., Damaševičius, R., Alouffi, B., and Alharbi, A. (2021). AI-Driven Framework for Recognition of Guava Plant Diseases through Machine Learning from DSLR Camera Sensor Based High Resolution Imagery. *Sensors*, 21(11), 3830. https://doi.org/10.3390/s21113830

15. Zheng, J., et al. (2025). Precision citrus segmentation and stem picking point localization using improved YOLOv8n-seg algorithm. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2025.1655093
16. Anonymous (2024). A Banana Disease Detection Using MobileNetV2 Model Based on Adam Optimizer. *Journal of Applied Informatics and Computing (JAIC)*, 8(1). https://jurnal.polibatam.ac.id/index.php/JAIC/article/download/9870/2910/32853
17. Anonymous (2025). Guava Disease Classification Using EfficientNet and Genetic Algorithm-Optimized XGBoost. *Journal of Intelligent Systems and Robotics*. https://shmpublisher.com/index.php/joiser/article/download/593/329/3906
18. A. K. Jain and N. Nain, "Mdtacnet: Mobilenet-densenet and transformer attention hybrid network for multi-crop and multi-disease classification," in *International Conference on Computer Analysis of Images and Patterns*, Springer, 2025, pp. 3–13.
19. A. K. Jain and N. Nain, "Hybrid Deep Learning Framework for Cashew Leaf Disease Classification Using Model Concatenation," *Technoarete Transactions on Advances in Computer Applications (TTACA)*, vol. 4, no. 2, pp. 14–18, 2025.
20. A. K. Jain, K. Harshavardhan, and N. Nain, "Multi crop Chilli and Rice Diseases Classification using Multi Concatenate of Models," in *2025 12th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, IEEE, 2025, pp. 416–420.
21. A. K. Jain, D. Sharma, and N. Nain, "Disease Detection and Classification of Rice Leaves Using Transfer Learning with Concatenate," in *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, Springer, 2024, pp. 234–245.
22. A. K. Jain and N. Nain, "Attention-Enhanced Hybrid CNN Architecture for Multi-Crop and Multi-Disease Classification: A Case Study on Papaya and Sugarcane," in *10th International Conference on Computer Vision and Image Processing (CVIP 2025)*, Springer, 2025, pp. 20–31.
23. Jain, A. K., Nain, N., and Jain, A. (2026). ABCGMP Fruit and Leaf Disease Six Crop Dataset. *Mendeley Data*, V2. doi: 10.17632/fhbvmpcyy2.2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- history.txt
- llncsdoc.pdf
- readme.txt