

An Explainable Deep Learning Framework for Plant Leaf Disease Detection Using a Custom CNN Model

Sarah Shtawa
Computer Science Department
Libyan Academy, Gharyan, Libya
sarahshtawa111@gmail.com

Naser Alfed
Computer Science Department
Faculty of Science, Gharyan University,
Libya
naser.alfed@gu.edu.ly

Albahloul M Abood
Faculty of Information Technology
Gharyan University, Libya
Albahloul.abood@gu.edu.ly

Abstract— Agricultural sustainability relies heavily on the early detection of plant pathologies. However, manual diagnosis remains challenging even for experts. This study proposes a lightweight Custom Convolutional Neural Network (CNN) architecture for automated leaf disease detection. The model was evaluated against state-of-the-art frameworks, MobileNetV2 and EfficientNetB0, using a dataset of 15,649 images that integrates global data with locally sourced samples from Libya. To ensure robustness, k-Fold Cross-Validation was implemented under standardized conditions. The proposed Custom CNN achieved a competitive accuracy of 97.6%, closely matching EfficientNetB0 (98.4%). Despite the slight accuracy advantage of transfer learning models, the Custom CNN demonstrated superior computational efficiency and a significantly smaller architectural footprint. These results position the proposed model as an ideal candidate for deployment in resource-constrained environments and mobile-based diagnostic systems.

Keywords— Custom CNN, Deep Learning, EfficientNetB0, Explainable AI (XAI), Libyan Agriculture, MobileNetV2, Plant Leaf Diseases, Transfer Learning.

I. INTRODUCTION

Proactive crop health management has become a fundamental requirement in modern smart agriculture, particularly as plant diseases continue to threaten crop productivity, food security, and agricultural sustainability. Early detection of plant pathologies is essential for limiting disease spread, reducing economic losses, and supporting timely intervention strategies. However, manual diagnosis remains challenging because early symptoms are often subtle, visually similar, and difficult to distinguish even by experienced agricultural specialists. Traditional visual inspection is also labor intensive, subjective, and difficult to scale for large agricultural areas. [8]. Recent advances in deep learning have significantly improved automated plant disease detection by enabling models to extract discriminative visual features directly from leaf images. Convolutional Neural Networks have demonstrated strong performance in identifying complex pathological patterns, while transfer learning architectures such as MobileNetV2 and EfficientNetB0 have further improved classification accuracy by leveraging knowledge from large scale image datasets. Nevertheless, high accuracy alone is not sufficient for practical agricultural deployment. Real world diagnostic

systems, especially those intended for mobile or edge environments, require models that are not only accurate, but also lightweight, computationally efficient, and interpretable.

A. Motivation and Research Gap

Despite the significant progress achieved in plant leaf disease detection, several important gaps remain insufficiently addressed. First, many existing studies rely heavily on transfer learning architectures that achieve high predictive performance but often require greater computational resources, making them less suitable for deployment on mobile devices or resource constrained edge platforms. Second, most publicly available datasets are derived from controlled or globally collected repositories, while limited attention has been given to locally acquired agricultural data from specific regional environments such as Libya. This limits the ability of existing models to reflect local field conditions, natural lighting variations, and region specific visual characteristics.

Moreover, many deep learning based diagnostic systems operate as black box models, offering limited insight into the visual evidence that supports their predictions. This lack of interpretability can reduce user trust, particularly in agricultural decision making where incorrect diagnosis may lead to unnecessary treatment or crop loss. Therefore, there remains a clear need for a lightweight, explainable, and regionally relevant deep learning framework capable of achieving high diagnostic performance while maintaining practical suitability for real world agricultural applications.

B. Contributions

To address these limitations, this study proposes an explainable deep learning framework for plant leaf disease detection based on a lightweight Custom CNN architecture. The proposed model is evaluated against two well established transfer learning architectures, MobileNetV2 and EfficientNetB0, using a hybrid dataset that combines PlantVillage images with locally collected Libyan agricultural samples. The main contributions of this work are summarized as follows:

A lightweight Custom CNN architecture is designed to achieve a practical balance between classification accuracy

and computational efficiency for resource constrained deployment.

A hybrid dataset is constructed by integrating global PlantVillage images with locally collected Libyan samples, improving the regional relevance of the diagnostic framework.

A comparative evaluation is conducted against MobileNetV2 and EfficientNetB0 using consistent preprocessing, validation, and testing conditions.

Explainable AI techniques, namely Grad-CAM and Score-CAM, are incorporated to visualize model attention and improve the transparency of diagnostic decisions.

The computational efficiency of the evaluated models is analyzed in terms of parameter count, model size, and inference time to assess their suitability for mobile and edge based agricultural applications.

C. Paper Organization

The remainder of this paper is organized as follows. Section II reviews related studies on machine learning and deep learning approaches for plant disease detection. Section III presents the proposed methodology, including dataset acquisition, preprocessing, model architecture, training strategy, and evaluation metrics. Section IV discusses the experimental results and comparative performance of the evaluated models. Section V summarizes the main findings. Section VI presents the interpretability analysis using Grad-CAM and Score-CAM. Finally, Section VII concludes the paper and outlines the main limitations and future research directions.

II. LITERATURE REVIEW

The evolution of automated plant disease diagnosis has transitioned from traditional image processing to sophisticated Deep Learning architectures. While early methods relied on manual feature engineering, the introduction of Convolutional Neural Networks (CNNs) has significantly improved classification accuracy across various botanical datasets.

The following previous studies are presented from the earliest to the latest ones:

A. Machine learning Techniques

Mohammed and Amel [1] conducted an investigation in 2019 into the efficacy of traditional algorithms by developing a framework centered on the Support Vector Machine (SVM) classifier. Their methodology incorporated rigorous preprocessing stages, such as fuzzy histogram equalization, to refine visual contrast before manually extracting color and texture descriptors from a 799-sample dataset. Although the study achieved a classification accuracy of approximately 88.1%, the authors emphasized that reliance on manual feature engineering remains a significant technical constraint, particularly for deployment in unpredictable real-world environments.

Mona et al. [2] introduced a hybrid methodology in 2023 for the classification of botanical infections by integrating improved K-Means clustering with a Support Vector Machine (SVM) classifier. Utilizing the Plant Village dataset, the investigators implemented a sophisticated pipeline of preprocessing operations, including noise suppression, contrast optimization, and color space conversion to facilitate precise disease region segmentation. While this integrated framework achieved a remarkable classification accuracy of approximately 97%, the authors identified a critical vulnerability in the reliance on manual feature engineering. Specifically, the study noted that such approaches exhibit limited adaptability when exposed to diverse or unfamiliar datasets. Consequently, the researchers recommended the adoption of deep learning architectures as a more robust alternative, citing their inherent capacity for autonomous feature discovery and superior generalization.

Somya et al. [3] conducted a comparative evaluation in 2025 of various supervised learning algorithms to determine their efficacy in identifying botanical pathologies across five distinct categories within the Plant Village dataset. The study assessed a wide array of classifiers, including SVM, Random Forest, Naïve Bayes, Decision Tree, k-Nearest Neighbor, and Multilayer Perceptron. Central to their methodology was the implementation of rigorous preprocessing protocols to refine image quality for improved classification outcomes. Their experimental results indicated that the Random Forest algorithm attained the highest performance with an accuracy of approximately 91%. Ultimately, the authors underscored the necessity of dataset diversity for enhancing model stability and recommended a direct comparison between these traditional architectures and Deep Learning (DL) frameworks, which possess the capacity for autonomous feature extraction.

Utkarsha et al. [4] conducted a comparative investigation in 2020 to evaluate the effectiveness of different diagnostic frameworks for botanical pathologies. Their methodology utilized a traditional pipeline involving image segmentation and texture analysis via the Gray-Level Co-occurrence Matrix (GLCM), which was subsequently classified using an SVM. This conventional approach was benchmarked against a Convolutional Neural Network (CNN) architecture. The experimental findings revealed that the DL framework significantly surpassed the traditional SVM model, a result attributed to the CNN's inherent capacity for autonomous feature acquisition from raw visual data. Furthermore, the study demonstrated that integrating data augmentation techniques substantially bolstered model performance by diversifying the training distribution.

B. Deep learning Techniques

Rauf et al. [5] explored in 2025 the efficacy of various computational frameworks for plant pathology identification by benchmarking a custom CNN against pre-trained models on 20,000 images. While the specialized CNN attained 87.5% accuracy, ResNet50 achieved a superior precision of 94.8%. Despite this gap, the researchers noted that the simpler model required fewer computational resources, suggesting its

viability for limited hardware environments. The study concluded by advocating for architectural optimizations to harmonize the relationship between diagnostic accuracy and operational efficiency.

Khalifa et al. [6] developed a dedicated neural architecture in 2025 to mitigate the impact of plant diseases on yields. Their sequential design comprised three convolutional layers integrated with dimensionality reduction and dropout mechanisms to prevent over fitting. By utilizing the ReLU activation function and a Softmax layer for multiclass categorization, the model achieved an accuracy of 96.54% on 5,000 images. The authors posited that such tailored frameworks provide a robust baseline for comparison with advanced pre-trained networks, highlighting the potential of specialized deep learning in precision agriculture.

Oni and Parma [7] developed a tailored CNN in 2025 for real-time plant disease detection, motivated by significant crop yield losses. The proposed network was benchmarked against YOLOv5, MobileNetV2, and ResNet18 using a local dataset from Bangladesh. The custom model achieved an accuracy of 95.2%, outperforming the transfer learning approaches. Their findings suggest that task-specific architectures can surpass pre-trained models when designed for particular environmental conditions. However, the study recommended expanding the dataset diversity to ensure robustness under varied scenarios.

Natarajan et al. [8] utilized DL methodologies in 2020 to establish an automated framework for diagnosing tomato infections. Their setup compared a custom CNN against a ResNet50 architecture using 1,090 real-field images capturing various disease stages. The results confirmed that CNN models are proficient in identifying the visual signatures of plant pathologies, ensuring consistent classification. Ultimately, the researchers highlighted that expanding the data repository and incorporating broader disease varieties are crucial for enhancing the robustness of DL systems in actual agricultural environments.

Raigonda and Angali [9] developed a DL methodology in 2025 for diagnosing rice foliage pathologies. Their research benchmarked a basic CNN against a VGG16-based transfer learning (TL) model on 1,800 images. Results indicated that TL significantly outperformed the custom CNN, achieving 95.60% accuracy compared to 85.62%. The researchers emphasized the efficiency of pre-trained networks for limited datasets and recommended exploring modern architectures like EfficientNet and MobileNet to optimize performance, alongside tools like Grad-CAM for enhanced reliability.

Table I: Summary of Related Work and Identified Research Gaps

Reference	model	Dataset	Accuracy	Limitations (The Gap)
Rauf et al. (2025) [5]	Custom CNN vs. ResNet50	PlantVillage (Simple Split Dataset)	94.8% (ResNet50)	High computational demand for ResNet50; Custom CNN accuracy was

				relatively low (87.5%).
Khalifa et al. (2025) [6]	Custom CNN	PlantVillage (Simple Split Dataset)	96.54 %	Lacked benchmarking against diverse pre-trained models and XAI transparency.
Oni and Parma. (2025) [7]	Custom CNN vs. YOLOv5	Local (Bangladesh) (Simple Split Dataset)	95.2%	Recommended expanding dataset diversity and environmental robustness.
Raigonda et al. (2025) [9]	Custom CNN vs. VGG16	plant Village (Simple Split Dataset)	95.6% (VGG16)	VGG16 is heavy for mobile use; recommended exploring XAI tools like Grad-CAM.
This Study. (2026)	Proposed Custom CNN Vs. MobileNet V2, EfficientNetB0	(Hybrid : Global + Libya) (5-Fold CV)	Custom CNN 97.6%	Optimized for mobile use; integrates XAI (Grad-CAM/Score-CAM) and local Libyan data.

As synthesized in Table I, while recent state-of-the-art models have achieved impressive accuracy, a persistent challenge remains in balancing high performance with computational frugality and model transparency. Most current frameworks either rely on heavy transfer learning architectures that are unsuitable for real-time mobile deployment or lack the interpretability required for agricultural confidence. Furthermore, the reliance on idealized datasets often limits their robustness in diverse regional climates. Our proposed framework addresses these specific gaps by introducing a lightweight Custom CNN that ensures high diagnostic precision (97.6%) while maintaining architectural simplicity, validated through 5-fold cross-validation and enhanced with Explainable AI for visual transparency. By integrating local Libyan field data, this study also bridges the geographical gap often found in current plant pathology research.

III. PROPOSED METHODOLOGY

This section delineates the systematic approach adopted to develop an efficient diagnostic framework for plant leaf

diseases. The methodology encompasses three primary phases: the acquisition and curation of the botanical dataset, the implementation of essential preprocessing and augmentation techniques, and finally, the design and training of the proposed deep learning architectures optimized for resource-constrained environments.

A. Dataset Acquisition

The experimental foundation of this research is a hybrid dataset comprising 15,649 images, structured to ensure both diversity and local relevance. This collection integrates two primary sources:

1. Global Repository: A subset of images was curated from the PlantVillage repository (via Kaggle), <https://www.kaggle.com/datasets/moazeldsokyx/plantvillage> covering eight distinct classes: Pepper (Bacterial Spot, Healthy), Potato (Early Blight, Late Blight, Healthy), and Tomato (Bacterial Spot, Yellow Leaf Curl Virus, Leaf Spot, Healthy).
2. Local field data: To enhance the model's practical reliability in specific agricultural contexts, original samples were collected from the western region, specifically from the city of Gharyan. This component includes local datasets for early blight on pears and for healthy peppers.

Table II: Description of the dataset used

Plant Type	Total number for each category	Data Source
bacterial spots on peppers	1305	PlantVillage
healthy peppers	1269	Local data
early potato blight	1661	PlantVillage
late potato blight	1827	PlantVillage
healthy potatoes	1527	PlantVillage
bacterial spots on tomatoes	1272	PlantVillage
Tomato septoria leaf spot	2149	PlantVillage
yellow leaf curl virus on tomatoes	2943	PlantVillage
healthy tomatoes	1598	PlantVillage
Pear Early blight	98	Local data

As detailed in Table II, this study presents a unique hybrid dataset that bridges the gap between controlled environments and real-world agricultural conditions. A key contribution is the inclusion of locally sourced imagery from a Libyan farm, captured on mobile devices under diverse natural lighting and backgrounds. Unlike standard databases, this local integration enhances the model's environmental robustness and regional adaptability, ensuring that the framework is specifically tailored to the unique variability and visual noise prevalent in North African agricultural ecosystems. This local dataset serves as a crucial benchmark for translating deep learning models from theoretical validation to practical, field-based diagnostic application.

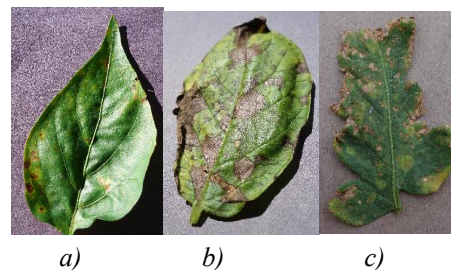


Fig.1. Representative Samples from the available dataset (PlantVillage): (a) Bacterial spots on peppers, (b) potatoes infected with early blight, (c) bacterial spots on tomatoes

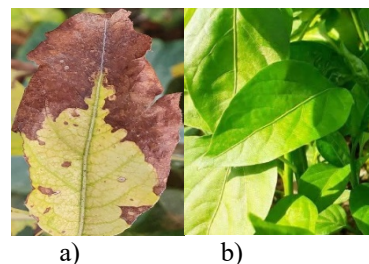


Fig.2. Representative samples from the local dataset: (a) early blight in pears, (b) healthy bell peppers

B. Data Partitioning and Validation Strategy

To ensure model generalizability and mitigate selection bias, a robust multi-stage validation strategy was implemented the primary dataset was subjected to a K-fold cross-validation protocol ($K=5$), where in each iteration, $K-1$ segments (80%) were utilized for training while one segment (20%) was reserved for validation. This mechanism ensures that every data point serves for both training and verification across all iterations, with samples shuffled in each batch to eliminate bias [17]. Upon completing all folds, the aggregate performance was average to provide a stable assessment and minimize the impact of potential outliers. Finally, an independent 15% portion of the original data was withheld for final testing; these samples remained separate and were not subjected to any data augmentation techniques to ensure

an objective evaluation of the model's performance in real-world scenarios [17].

C. Image Preprocessing and Augmentation

To enhance image quality and ensure uniform input for the deep learning models, a standardized preprocessing pipeline was implemented. All images were resized to a fixed resolution of 224 times 224 pixels, a dimension selected to balance the preservation of critical visual features with computational efficiency [17]. To mitigate overfitting and improve model robustness, a dynamic data augmentation strategy was integrated into the training phase using the ImageDataGenerator library. This process involved applying various transformations to the training samples, allowing the models to generalize across a more diverse distribution of visual data. Crucially, these augmentation techniques were exclusively applied to the training partitions; the validation and testing sets remained in their original state to ensure an unbiased assessment of the model's reliability and its capacity for generalization in real-world scenarios.

D. Proposed Custom CNN Architecture

The proposed CNN framework was engineered using a sequential arrangement of functional layers, specifically optimized for the complexities of botanical disease classification. The structural components of the architecture are detailed as follows:

1. **Input and Feature Extraction Backbone:** The network accepts RGB color images with dimensions of 224 x 224 x 3. The architecture incorporates four successive convolutional blocks with an incrementally expanding filter depth from 32 to 64, 128, and 256. Each convolution utilizes the Rectified Linear Unit (ReLU) activation function to introduce non-linearity, enabling the model to learn complex hierarchical representations.
2. **Normalization and Spatial Reduction:** Following each convolutional operation, a Batch Normalization layer is integrated to stabilize the internal covariate shift, thereby accelerating convergence and enhancing training stability. Subsequently, a 2 x 2 MaxPooling layer is applied to down sample the feature maps, reducing spatial dimensionality while preserving the most prominent visual characteristics.
3. **Global Feature Aggregation and Regularization:** To transition to the classification stage while preventing parameter explosion, Global Average Pooling (GAP) is utilized instead of traditional flattening. This substantially reduces the total number of trainable parameters, mitigating overfitting. The classification head consists of a Dense layer with 256 neurons (ReLU-activated), followed by a Dropout layer (rate = 0.5) to bolster robustness by preventing over-reliance on specific neural pathways.
4. **Output Stage:** The final layer comprises 10 neurons, corresponding to the target classes. A Softmax

function is applied to transform raw outputs into a probability distribution, facilitating precise multiclass categorization.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
batch_normalization (BatchNormalization)	(None, 222, 222, 32)	128
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18,496
batch_normalization_1 (BatchNormalization)	(None, 109, 109, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_2 (Conv2D)	(None, 52, 52, 128)	73,856
batch_normalization_2 (BatchNormalization)	(None, 52, 52, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 26, 26, 128)	0
conv2d_3 (Conv2D)	(None, 24, 24, 256)	295,168
batch_normalization_3 (BatchNormalization)	(None, 24, 24, 256)	1,024
max_pooling2d_3 (MaxPooling2D)	(None, 12, 12, 256)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 256)	0
dense (Dense)	(None, 256)	65,792
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 10)	2,570

Fig.3 Structure of the proposed model

E. Comparative models

To rigorously validate the efficacy of the proposed Custom CNN, a comparative analysis was conducted against two state-of-the-art architectures: EfficientNetB0 and MobileNetV2. To ensure an equitable evaluation, all models utilized the identical hybrid dataset, 5-fold cross-validation, and preprocessing pipeline. Both models were implemented using a four-stage Transfer Learning protocol:

1. **Architectural Initialization:** Pre-trained frameworks (EfficientNetB0 and MobileNetV2) were loaded with ImageNet weights. The original top classification layers were excluded to facilitate the integration of a customized head tailored to the 10 botanical classes [20].
2. **Feature Preservation (Freezing):** The foundational backbone layers of both networks were frozen. This strategic constraint preserves the generalized visual descriptors learned from large-scale datasets while significantly reducing computational overhead and the risk of overfitting [24].
3. **Custom Head Integration:** A specialized classification head was appended to each model, comprising a Global Average Pooling (GAP) layer for spatial feature aggregation, a Dense layer for high-level representation, and a Dropout mechanism (0.5) to enhance regularization.
4. **Targeted Optimization:** Training was restricted solely to the newly integrated layers. This approach accelerates convergence by adapting the network to the specific characteristics of the plant pathology dataset while maintaining consistency with the hyper-parameters used in the Custom CNN.

EfficientNetB0: Selected for its optimal balance of depth and width via Compound Scaling, ensuring high precision with fewer parameters [20].

MobileNetV2: Engineered with depth-wise separable convolutions, making it ideal for resource-constrained agricultural environments in Libya due to its minimal computational complexity [24].

F. Model Training and Optimization Strategy

The training phase was implemented using the TensorFlow/Keras library, with a suite of optimization techniques configured to ensure stable convergence and high diagnostic precision. The Adam (Adaptive Moment Estimation) optimizer was employed to update network weights by integrating momentum-based optimization with adaptive learning rates [15]. Performance was monitored via Sparse Categorical Cross-Entropy loss and overall classification Accuracy.

To manage computational resources and ensure gradient stability, the dataset was processed in a Batch Size of 32 samples per step. The training was initially set for 20 epochs; however, a dynamic supervision strategy was applied to maximize generalization:

- **Early Stopping:** A safeguard mechanism was implemented to monitor validation loss. Training was pre-emptively concluded if no quantitative improvement was observed for a patience window of five consecutive epochs, with the model reverting to the state yielding the minimum validation loss [17].
- **Learning Rate Reduction (ReduceLROnPlateau):** To refine the optimization path, the learning rate was automatically scaled down by 20% if the loss reduction plateaued for three epochs, facilitating more granular weight updates [17].
- **Model Checkpoint:** This ensured the best version of the model representing the peak of validation performance was preserved for final evaluation.

G. Performance Evaluation and Interpretability

To provide a comprehensive assessment of the diagnostic framework, the models were evaluated using a multidimensional approach focusing on both statistical reliability and practical transparency:

1. **Statistical Metrics:** Beyond global Accuracy, the models' proficiency was quantified using Precision, Recall, and the F1-Score. Additionally, a Confusion Matrix was utilized to visualize the distribution of predictions across the 10 botanical classes, ensuring that the model performs consistently across both healthy and pathological samples.
2. **Computational efficiency:** Given the goal of deploying these models in resource-limited environments, computational load was a key evaluation factor. This included measuring the inference response time per image, the total number of trainable parameters, and the model size to ensure an optimal balance between diagnostic accuracy and hardware requirements.

3. **Interpretable Artificial Intelligence (XAI):** To enhance the model's transparency and reliability, Grad-weighted class activation mapping techniques (Grad-CAM and Score-Cam) were implemented. These tools generate heat maps that highlight specific spatial regions within the leaf image that significantly influenced the model's classification decision. This ensures that the diagnostic process is based on relevant pathological features rather than irrelevant background noise.

IV. RESULTS AND DISCUSSION

This section presents a detailed account of the experimental results obtained from the evaluated constructs. The analysis begins by examining convergence patterns through training and validation paths. To ensure an objective assessment, performance was validated using an independent test set of 1822 unique units, which remained isolated throughout the development phase. Diagnostic capabilities were further examined using a multidimensional approach, including a confusion matrix, classification reports, and computational efficiency analysis.

A. Training and Validation Results

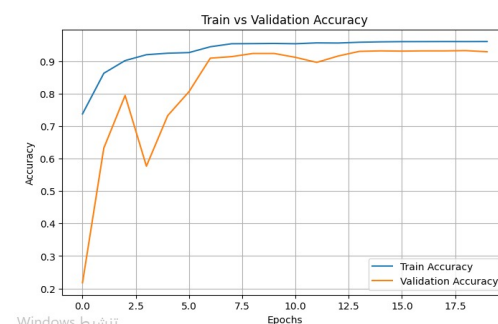


Fig.4 Train vs. validation Accuracy for a custom CNN model

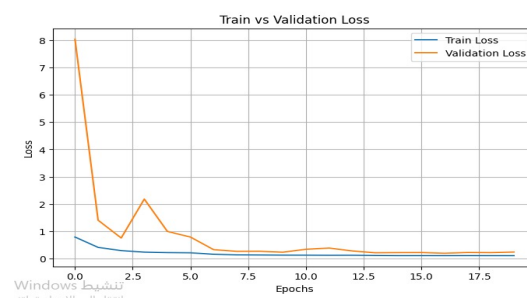


Fig.5 Train vs. Validation loss for a custom CNN model

The performance evolution in Figure 3 shows a consistent upward trend, with training accuracy starting at 73% and converging at 95%. Although the validation metric exhibited initial fluctuations common in multi-category classification of visually similar plant diseases it showed marked improvement after the sixth iteration. The loss trajectories in Figure 4 show a sharp initial decrease, indicating rapid improvement. The five-fold cross-validation achieved a robust average accuracy of 95% and a validation loss of

0.1409, confirming the structure's ability to generalize strongly and avoid over-allocation.

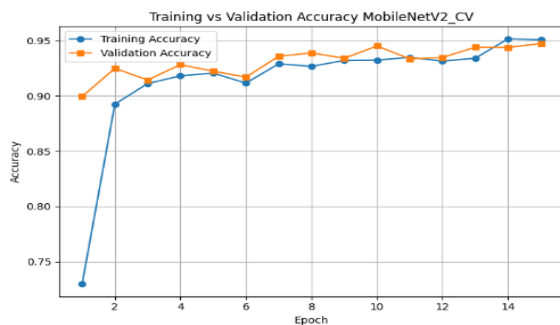


Fig.6. Train vs. Validation Accuracy for MobileNetV2 model



Fig.7. Train vs. Validation loss for MobileNetV2 model

As shown in Fig. 5, MobileNetV2 exhibited a rapid escalation in training accuracy. The narrow margin between training and validation trajectories indicates consistent stability. Under the cross-validation framework, the model attained a validation accuracy of 94.7%. The error reduction patterns 6 reached a stable state by the 15th epoch, recording a minimum validation loss of 0.1399.



Fig.8. Train vs. Validation Accuracy EfficientNetB0 model

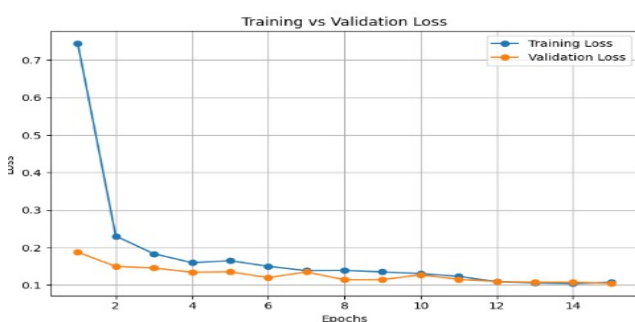


Fig.9 Train vs. Validation loss for EfficientNetB0 model

The architecture of the EfficientNetB0 model exhibited the highest convergence stability, as illustrated in Figures 7 and 8. The accuracy and loss curves stabilized at a steady state with minimal variation. EfficientNetB0 achieved an optimal verification accuracy of 96.2% and a loss of 0.1047, confirming its exceptional predictive power.

The comparative data highlights that EfficientNetB0 emerged as the top-performing model in terms of raw accuracy. However, the Proposed Custom CNN secured the second-highest accuracy (95%), notably outperforming MobileNetV2. This is a significant finding, as it demonstrates that a specifically engineered, lighter architecture can rival or exceed the performance of established transfer learning models while maintaining lower complexity. While validation results are promising, the following sections will analyze the models' performance on the independent test dataset to confirm their reliability in real-world diagnostic scenarios.

B. Performance Evaluation on Independent Testing

Following the optimization and validation stages, the predictive reliability of the three architectures was scrutinized using an independent testing set of 1,822 samples. This isolated dataset provides an objective measure of the models' efficacy in generalizing to novel data in real-world agricultural scenarios.

57/57	46s 792ms/step	precision	recall	f1-score	support
Pepper__bell__Bacterial_spot	0.9930	0.9792	0.9860	144	
Pepper__bell__healthy	0.9927	0.9855	0.9891	138	
Potato__Early_blight	0.9476	0.9945	0.9705	182	
Potato__Late_blight	1.0000	0.9415	0.9698	205	
Potato__healthy	0.9816	1.0000	0.9907	213	
Tomato__Bacterial_spot	0.9857	0.8571	0.9169	161	
Tomato__Septoria_leaf_spot	0.9004	1.0000	0.9476	226	
Tomato__Tomato_YellowLeaf__Curl_Virus	1.0000	0.9942	0.9971	345	
Tomato__healthy	1.0000	1.0000	1.0000	189	
pear__Early_blight	1.0000	1.0000	1.0000	19	
accuracy			0.9764	1822	
macro avg	0.9801	0.9752	0.9768	1822	
weighted avg	0.9779	0.9764	0.9763	1822	

Fig.10 Results of the CNN-Custom Architecture Classification Report.

The classification performance of the Custom CNN demonstrates exceptional stability, achieving a weighted F1-average score of 97.63% and an overall accuracy of 97.68%. This indicates a highly balanced diagnostic capability across the 10 botanical categories. Notably, the model achieved a 100% Recall rate for several critical classes, including Healthy Tomato, Healthy Potato, Septoria Leaf Spot, Yellow Leaf Curl Virus, and Pear Early Blight, signifying zero misclassification for these groups. A minor diagnostic challenge was observed in the Tomato Bacterial Spot category, where the recall dipped to 85.71% due to its high visual similarity with other pathological spotting patterns.

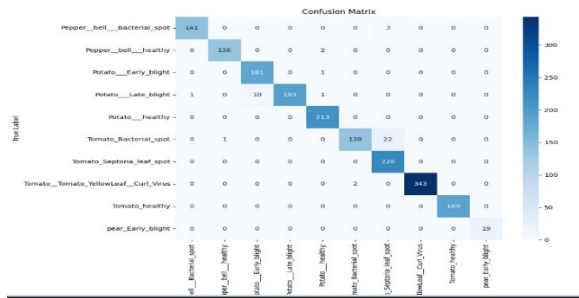


Fig.11 Confusion matrix for a custom CNN Architecture.

The confusion matrix highlights the architecture's robust predictive power, with high numerical values along the principal diagonal confirming accurate classification across most categories. While the majority of classes achieved full accuracy, a minor diagnostic overlap was observed: 22 instances of Tomato Bacterial Spot were misclassified as Tomato Septoria Leaf Spot. This specific confusion is attributed to the high visual resemblance in foliar spotting patterns between the two diseases. Conversely, the model achieved perfect recognition for all Septoria Leaf Spot samples, demonstrating its proficiency in distinguishing complex botanical features in a multi-class environment.

	precision	recall	f1-score	support
Pepper_bell_Bacterial_spot	0.9930	0.9861	0.9895	144
Pepper_bell_healthy	0.9787	1.0000	0.9892	138
Potato_Early_blight	0.9891	0.9945	0.9918	182
Potato_Late_blight	0.9901	0.9895	0.9893	205
Potato_healthy	0.9953	0.9906	0.9929	213
Tomato_Bacterial_spot	0.9496	0.8199	0.8800	161
Tomato_Septoria_leaf_spot	0.8884	0.9513	0.9188	226
Tomato_Tomato_YellowLeaf_Curl_Virus	0.9971	0.9942	0.9956	345
Tomato_healthy	0.9643	1.0000	0.9818	189
pear_Early_blight	1.0000	1.0000	1.0000	19
accuracy			0.9720	1822
macro avg	0.9746	0.9717	0.9725	1822
weighted avg	0.9725	0.9720	0.9717	1822

Fig.12. Results of the MobileNetV2 Architecture Classification Report.

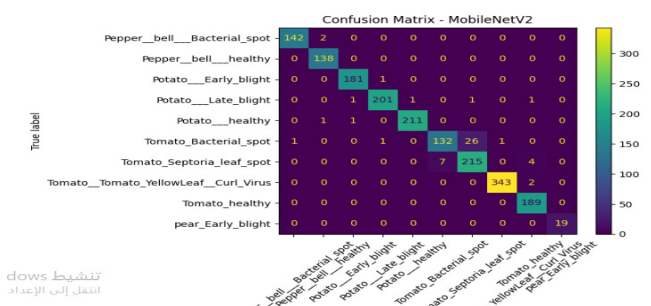


Fig.13 Confusion matrix results for the MobileNetV2 Architecture

The MobileNetV2 model demonstrated high classification efficiency with a test accuracy of 97.75% and an F1 score of 97.20%, indicating strong generalizability across most healthy and diseased categories. However, a detailed analysis of the classification report and confusion matrix (Figure 11,12) reveals a significant diagnostic challenge in the bacterial tomato spot category, where it scored 81.93%. The

model exhibited two-way confusion with Septoria leaf spot, misclassifying 26 cases of bacterial spot and 7 cases of Septoria. This mutual ambiguity, resulting from the overlap of visual features of these two diseases, suggests that MobileNetV2 had greater difficulty distinguishing features compared to the proposed custom CNN, which maintained higher sensitivity in these specific categories.

==== Test Classification Report =====				
	precision	recall	f1-score	support
Pepper_bell_Bacterial_spot	1.0000	0.9792	0.9895	144
Pepper_bell_healthy	0.9787	1.0000	0.9892	138
Potato_Early_blight	1.0000	1.0000	1.0000	182
Potato_Late_blight	0.9903	0.9951	0.9927	205
Potato_healthy	0.9953	0.9953	0.9953	213
Tomato_Bacterial_spot	0.9456	0.8634	0.9026	161
Tomato_Septoria_leaf_spot	0.9121	0.9646	0.9376	226
Tomato_Tomato_YellowLeaf_Curl_Virus	1.0000	1.0000	1.0000	345
Tomato_healthy	1.0000	1.0000	1.0000	189
pear_Early_blight	1.0000	1.0000	1.0000	19
accuracy			0.9808	1822
macro avg	0.9822	0.9798	0.9807	1822
weighted avg	0.9810	0.9808	0.9806	1822

Fig.14 Results of the EfficientNetB0 Architecture Classification Report.

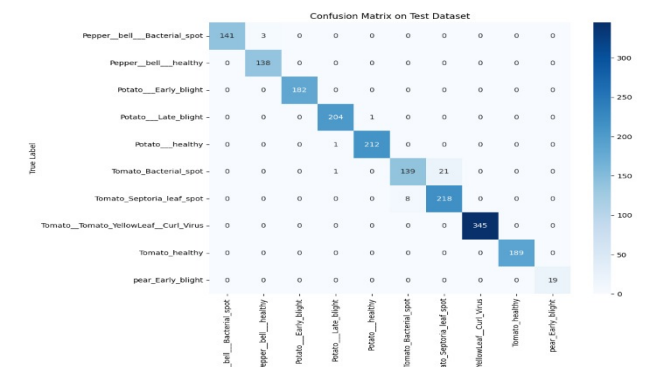


Fig. 15 Confusion matrix results for the EfficientNetB0 Architecture.

The EfficientNetB0 architecture demonstrated the highest statistical precision among the evaluated models, achieving a weighted F1-score of 98.86%. As illustrated in the classification report (Fig.13) and confusion matrix (Fig. 14), the model attained a perfect 100% success rate across the majority of categories, including Healthy Bell Pepper and Yellow Leaf Curl Virus. While the Tomato Bacterial Spot remained the most challenging class, EfficientNetB0 exhibited superior feature extraction by limiting misclassifications to only 21 instances. This refined ability to distinguish complex pathological patterns confirms the efficacy of Compound Scaling in transfer learning; however, this high accuracy is coupled with a higher computational complexity compared to the proposed Custom CNN.

Table III: Comparison of the performance of the models evaluated on the independent testing group

Model	Accuray	Loss	Test accuray	Test loss	Macro F1-score
Custom CNN	0.9768	0.0453	0.9764	0.0969	0.9768
Mobile Net-V2	0.9775	0.0581	0.9720	0.0796	0.9725
EfficientNetB0	0.9844	0.1109	0.9808	0.0681	0.9807

Table IV: Results of the computational efficiency of the models

Model	Number of total parameters	Model size(MB)	Inference time when testing a sample (seconds)
Custom CNN	458,698	5.33 MB	0.51 s
MobileNetV2	2,588,492	12.91MB	80.56 s
EfficientNetB0	4,380,077	19.93MB	46.90 s

V. SUMMARY OF THE RESULTS

The experimental evaluation reveals that while EfficientNetB0 achieved the highest classification accuracy (98.44%), it recorded the highest computational cost in terms of model size and inference time. In contrast, the Proposed Custom CNN demonstrated an optimal trade-off, achieving a highly competitive classification accuracy of 97.68% while maintaining the lowest resource requirements across all efficiency metrics. A common diagnostic challenge was observed across all models in distinguishing between Tomato Bacterial Spot and Septoria Leaf Spot due to high visual similarity; however, the Custom CNN showed superior sensitivity in isolating the unique histological features of Septoria compared to MobileNetV2. Ultimately, while EfficientNetB0 is suitable for high-resource cloud applications, the Proposed Custom CNN emerges as the most viable and balanced solution for real-time diagnostics in resource-constrained environments, such as mobile deployment in local Libyan farms.

VI. INTERPRETATION OF MODELS DECISIONS

This section establishes a framework for elucidating the internal logic of the assessed structures using advanced visual interpretation methodologies. The primary objective is to identify the specific spatial features and areas of interest prioritized by the models during the classification process.

A significant portion of this analysis is dedicated to examining the more complex categories, particularly those where the models made inaccurate predictions. By analyzing these cases alongside successfully identified samples, the study provides a diagnostic explanation for misclassification among visually similar diseases. This comparative approach facilitates a deeper understanding of the patterns that promote correct behavior versus those that lead to errors in the Custom CNN, MobileNetV2, and EfficientNetB0 frameworks.

A. Interpreting Custom Model decisions Using Score-CAM

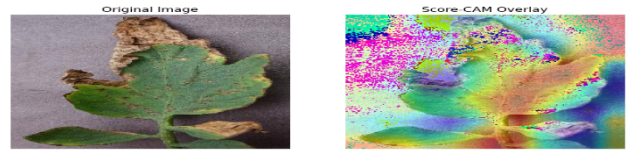


Fig.16. Analysis of a misclassification case of the Tomato Bacterial Spot category for a custom CNN model

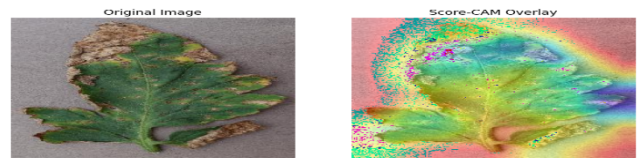


Fig.17. A correct classification analysis for the Tomato Bacterial Spot category for a custom CNN model

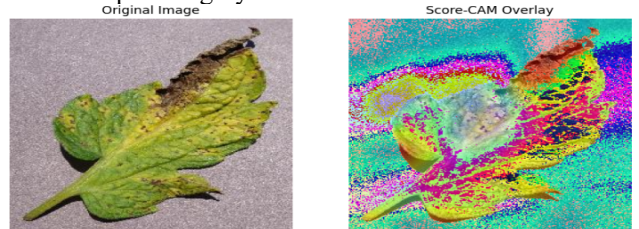


Fig.18. Analysis of a correct classification case for the Tomato Septoria Leaf Spot category for a custom CNN model

B. Interpreting MobileNetV2 decisions using Grad-CAM

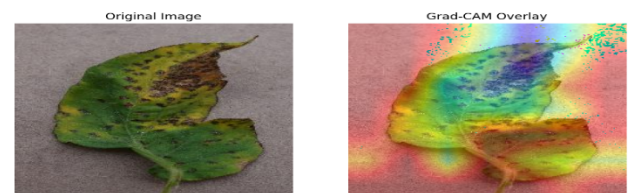


Fig.19. Analysis of a misclassification case of the Tomato Bacterial Spot category for a MobileNetV2

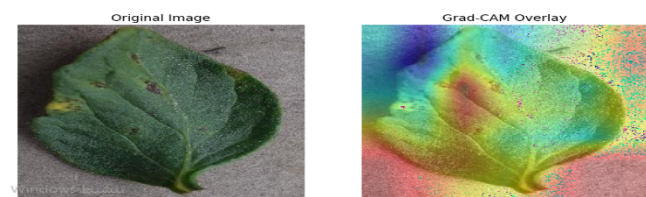


Fig.20. A correct classification analysis for the Tomato Bacterial Spot category for a MobileNetV2



Fig.21. Analysis of a correct classification case for the Tomato Septoria Leaf Spot category for a MobileNetV2

C. Interpreting EfficientNetB0 decisions Using Grad-CAM



Fig.22. Analysis of a misclassification case of the Tomato Bacterial Spot category for an EfficientNetB0

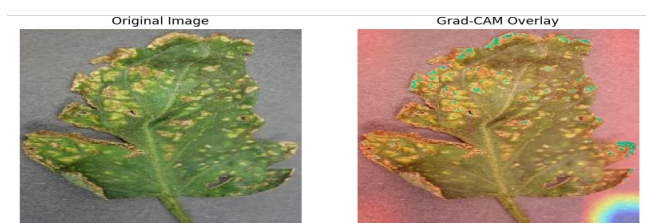


Fig.23. correct classification analysis for the Tomato Bacterial Spot category for an EfficientNetB0

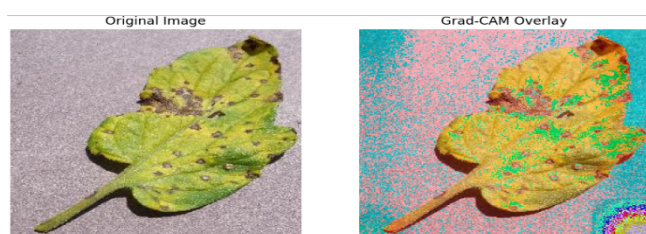


Fig.24. Analysis of a correct classification case for the Tomato Septoria Leaf Spot category for an EfficientNetB0

The visual interpretability analysis provides a profound understanding of the decision-making logic across the

evaluated architectures. The findings are synthesized as follows:

- Custom Convolutional Neural Network (High Specificity): Score-CAM graphs show that the proposed custom convolutional neural network maintains a precise diagnostic focus on actual pathological patterns, with minimal reliance on broader image context or background distortions. This confirms that the custom architecture has successfully learned to isolate disease-specific indicators. The occasional confusion between bacterial spot and septoria leaf spot is attributed to the visual similarity of these symptoms and not to weaknesses in the architecture, highlighting its reliability in feature-limited environments.
- MobileNetV2 (General focus versus detailed focus): In contrast, MobileNetV2 exhibited a broader and less granular visual focus. The heat maps reveal that this model tends to incorporate the entire leaf structure and background elements into its decision-making. This reliance on global patterns explains its higher frequency of binary confusion; it lacks the fine-grained sensitivity required to distinguish between diseases with highly similar textural characteristics, despite its established transfer learning weights.
- EfficientNetB0 (Superior Precision): The Grad-CAM results for EfficientNetB0 underscore its superior precision in identifying subtle pathological nuances. The model demonstrated the most consistent and localized activations, focusing strictly on the most distinctive disease patterns. This interpretative accuracy directly correlates with its optimal performance in the confusion matrix, confirming that its high classification success is rooted in a robust and stable understanding of complex botanical features.

VII. CONCLUSION AND FUTURE WORK

This research evaluated the effectiveness of deep learning methodologies for plant leaf disease classification by benchmarking the proposed Custom CNN against MobileNetV2 and EfficientNetB0. By integrating the PlantVillage dataset with locally collected samples from Libyan agricultural environments, the study provided a more practical basis for evaluating model performance under both standard and regionally relevant conditions.

The empirical results highlight a clear trade off between predictive accuracy and computational efficiency. EfficientNetB0 achieved the highest classification accuracy of 98.4%, confirming its strong feature extraction capability. However, the proposed Custom CNN achieved highly competitive performance while maintaining the lowest computational overhead in terms of parameter count, model size, and inference cost. This makes the Custom CNN a more practical candidate for resource constrained environments, including mobile based diagnostic systems and edge computing applications in local farms.

The integration of Explainable AI techniques, namely Grad-CAM and Score-CAM, further enhanced the transparency of the diagnostic process. The visual interpretation results confirmed that models with more localized attention to symptomatic regions produced more reliable decisions, particularly when distinguishing visually similar diseases such as Tomato Bacterial Spot and Septoria Leaf Spot. These findings demonstrate that a carefully designed lightweight CNN can provide a balanced solution that combines accuracy, efficiency, and interpretability for real time agricultural applications.

Despite these promising results, the study has some limitations. First, the number of locally collected Libyan samples remains relatively limited compared with the PlantVillage subset. Second, the dataset covers a restricted number of plant species and disease categories, which may limit generalization to broader agricultural contexts. Third, although the proposed model was designed for resource constrained deployment, it has not yet been tested in a real mobile or edge computing environment.

Future work should therefore focus on expanding the local dataset by including more plant species, disease types, and images captured under diverse climatic and field conditions. Further research could also integrate advanced mechanisms such as attention modules, multi scale feature extraction, or Vision Transformers to improve discrimination between visually similar disease categories. Finally, deploying the proposed Custom CNN on mobile or edge devices would provide an important step toward validating its real time performance and practical usability in smart agriculture.

REFERENCES

- [1] S. Mohanty, D. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, p. 1419, Sep. 2016, doi: 10.3389/fpls.2016.01419.
- [2] M. Jamjoom, A. Elhadad, H. Abulkasim, and S. Abbas, "Plant Leaf Diseases Classification Using Improved K-Means Clustering and SVM Algorithm for Segmentation," *Computers, Materials & Continua*, vol. 76, no. 1, pp. 367–382, 2023, doi:10.32604/cmc.2023.037310.
- [3] S. R. Goyal, V. S. Kulkarni, R. Choudhary, and R. Jain, "A comparative analysis of efficacy of machine learning techniques for disease detection in some economically important crops," *Crop Protection*, vol. 190, Apr. 2025, Art. No. 107093, doi:10.1016/j.cropro.2024.107093.
- [4] U. N. Fulari, R. K. Shastri, and A. N. Fulari, "Leaf Disease Detection Using Machine Learning," *Journal of Seybold Report*, vol. 15, no. 9, p. 1828, Sep. 2020.
- [5] M. T. Rauf, M. A. Wazir, A. Khalid, D. Khan, and O. B. Samin, "Detecting Plant Leaf Diseases Using CNN Models: A Comparative Study," *Spectrum of Engineering Sciences*, vol. 3, no. 8, pp. 69–80, Aug. 2025.
- [6] A. Khalifa, K. Patel, S. Parmar, D. Patel, and J. B. Upadhyay, "Plant Disease Detection using a Deep Learning approach: a Custom CNN," *International Research Journal on Advanced Engineering Hub (IRJAEH)*, vol. 3, no. 10, pp. 3869–3876, Oct. 2025, doi:10.47392/IRJAEH.2025.0562.
- [7] M. K. Oni and T. T. Prama, "Optimized Custom CNN for Real-Time Tomato Leaf Disease Detection," *arXiv:2502.18521*, Feb. 2025.
- [8] V. A. Natarajan, M. Babitha, and M. S. Kumar, "Detection of disease in tomato plant using deep learning techniques," *International Journal of Modern Agriculture*, vol. 9, no. 4, pp. 525–540, 2020.
- [9] M. R. Raigonda and Anjali, "Identification and Classification of Rice Leaf Disease Using Hybrid Deep Learning," *Journal of Scientific Research and Technology*, vol. 3, no. 6, pp. 28–40, Jun. 2025.
- [10] N. Shelar, S. Shinde, S. Sawant, S. Dhumal, and K. Fakir, "Plant Disease Detection Using CNN," *ITM Web of Conferences*, vol. 44, pp. 03049, 2022, doi: 10.1051/itmconf/20224403049.
- [11] A. Bhargava, A. Shukla, O. P. Goswami, M. H. Alsharif, P. Uthansakul, and M. Uthansakul, "Plant Leaf Disease Detection, Classification, and Diagnosis Using Computer Vision and Artificial Intelligence: A Review," *IEEE Access*, vol. 12, pp. 37443–37469, 2024, doi: 10.1109/ACCESS.2024.3373001.
- [12] M. Abdullah, A. Hussain, S. Qadri, J. Jabeen, I. Hussain, M. Khan, and H. N. Faried, "Deep Learning-Based Rice Leaf Disease Detection Using Custom CNN Models," *Journal of Computing & Biomedical Informatics*, vol. 9, no. 02, Sep. 2025.
- [13] S. Patel, "Leaf Disease Detection Using Convolutional Neural Network," M.Tech. thesis, Dept. of Computer Engineering (Software Eng.), Birla Vishvakarma Mahavidyalaya (Engineering College), Gujarat Technological University, Vallabh Vidyanagar, Gujarat, India, Jul. 2020.
- [14] R. M. Alfayez, A. A.H. Hassan, and H. T. S. Alrikabi, "Machine Learning and Deep Learning Techniques: A Review," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 16, no. 4, pp. 174–184, Dec. 2024, doi: 10.29304/jqscsm.2024.16.41806.
- [15] X. Li et al., "Adam Reduces a Unique Form of Sharpness: Theoretical Insights Near the Minimizer Manifold," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Vancouver, Canada, 2025.
- [16] Tejaswini, P. Rastogia, S. Dua, Manikanta, V. Dagar, "Early Disease Detection in Plants using CNN", *ELSEVIER, Procedia Computer Science* 235 (2024) 3468–3478.
- [17] E. M. E. Rzaq, "Evaluation of Machine Learning Techniques for Offline Handwritten Arabic Signature Recognition," M.S. thesis,

Dept. of Computer Science, School of Basic Sciences, The Libyan Academy, Tripoli, Libya, Feb. 2025.

[18] N. B. Alshagi, Detection of Brain Tumor Using Convolution Neural Network Algorithm, M.S. thesis, *Dept. Electrical and Computer Engineering, Libyan Academy – School of Applied Sciences and Engineering, Tripoli, Libya*, 2025.

[19] H. Abukraa, S. G. Kumari, and F. Bshia, “Survey for legume and cereal viruses in Libya,” *Arab Journal of Plant Protection*, vol. 40, no. 3, pp. 222–230, 2022, doi:10.22268/AJPP-40.3.222230.

[20] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A Comprehensive Survey on Transfer Learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[21] A. Abughnia and F. Adam, “Onion Diseases in the Libyan Jamahiriya,” *The Libyan Journal of Agriculture*, vol. 8, no. 1, 1979. [Online]. Available: <https://journals.uot.edu.ly/index.php/ljagric/article/view/573>.

[22] F. J. P. Montalbo and A. S. Alon, “Empirical Analysis of a Fine-Tuned Deep Convolutional Model in Classifying and Detecting Malaria Parasites from Blood Smears,” *KSII Transactions on Internet and Information Systems*, vol. 15, no. 1, pp. 147–165, Jan. 2021, doi: 10.3837/tiis.2021.01.009.

[23] J. G. A. Barbedo, “Plant disease identification from individual lesions and spots using deep learning,” *Biosystems Engineering*, vol. 180, pp. 96–107, Apr. 2019, doi: 10.1016/j.biosystemseng.2019.02.002.

[24] R. Ribani and M. Marengoni, “A Survey of Transfer Learning for Convolutional Neural Networks,” in *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, São Paulo, Brazil, Oct. 2019, pp. 1–10, doi:10.1109/SIBGRAPI-T.2019.00010.

[25] M. C. Hinojosa Lee, J. Braet, and J. Springael, “Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores,” *Applied Sciences*, vol. 14, no. 21, Art. 9863, Oct. 2024. Doi: 10.3390/app14219863.