Potato Leaf Disease Classification Using an Efficient Convolutional Neural Network

SANDEEP MUHAL
[1] **Bikaner Technical University, Bikaner, Rajasthan, India**

Corresponding author: Sandeep Muhal (sandeepmuhal9982@gmail.com)

This work was supported by the Dr. Jai Bhaskar

ABSTRACT Early identification of potato leaf diseases plays a vital role in minimizing crop losses and enhancing agricultural productivity. Conventional disease diagnosis practices largely depend on manual inspection by experts, which is labor-intensive, subjective, and difficult to scale in large farming environments. Although recent advances in deep learning have enabled automated plant disease detection with high accuracy, many existing solutions rely on deep and computationally heavy architectures, resulting in slow training convergence and limited feasibility for deployment in resource-constrained agricultural settings.

This study presents a lightweight convolutional neural network specifically designed for efficient potato leaf disease classification with an emphasis on rapid training convergence. The proposed model is constructed using stacked convolutional layers integrated with batch normalization and ReLU activation to ensure stable feature learning. To reduce parameter complexity, global average pooling is employed in place of traditional fully connected layers. In addition, adaptive training strategies, including the Adam optimizer, dynamic learning rate adjustment, and early stopping, are utilized to improve optimization stability and accelerate convergence.

The effectiveness of the proposed approach is validated on a three-class potato leaf image dataset comprising Early Blight, Late Blight, and healthy leaf samples. Experimental evaluation demonstrates that the model achieves a classification accuracy of 98% within only 10 training epochs, while consistently maintaining strong precision, recall, and F1-score across all categories. Performance comparison with existing deep and transfer learning–based models highlights that the proposed architecture attains comparable or superior results with substantially reduced training time and model complexity.

The obtained results confirm that compact CNN architectures optimized for fast convergence offer a reliable and practical solution for real-time plant disease diagnosis, particularly in low-resource and smart agriculture applications.

INDEX TERMS Potato leaf disease classification, convolutional neural networks, lightweight CNN, fast convergence, plant disease detection, precision agriculture.

I. INTRODUCTION

Agriculture remains a fundamental pillar of global food security, with potato (*Solanum tuberosum*) ranked among the most extensively cultivated and consumed crops worldwide. Despite its economic and nutritional importance, potato production is highly vulnerable to several plant diseases, of which ***Early Blight*** and ***Late Blight*** are considered the most severe. These diseases can propagate rapidly across crop fields, leading to substantial yield reduction and economic losses when timely preventive measures are not adopted. As a result, early and accurate identification of potato leaf diseases is a critical requirement for sustainable agricultural productivity and effective crop management.

Conventional potato disease diagnosis primarily depends on visual examination conducted by experienced agricultural specialists. While this approach can provide reliable results in certain cases, it is inherently labor-intensive, subjective, and difficult to scale for large agricultural fields. Furthermore, in rural and resource-constrained regions, the availability of trained agronomists is often limited, which may lead to delayed diagnosis and improper treatment decisions. Such limitations significantly reduce the effectiveness of traditional disease management strategies and highlight the need for automated and intelligent diagnostic solutions.

In recent years, advancements in computer vision and artificial intelligence have enabled the development of data-driven approaches for agricultural disease detection. In particular, deep learning techniques—most notably ***Convolutional Neural Networks (CNNs)***—have demonstrated remarkable performance in image-based classification tasks. CNNs are capable of learning hierarchical feature representations directly from raw images, thereby eliminating the dependency on handcrafted feature extraction methods. This capability has led to significant improvements in plant disease classification accuracy under controlled experimental environments.

Multiple studies have reported that deep CNN architectures can achieve high recognition performance on benchmark plant disease datasets. However, these approaches often rely on deep and computationally intensive models such as VGG, ResNet, and other transfer learning–based frameworks. Although such architectures deliver impressive accuracy, they typically require large-scale annotated datasets, prolonged training durations, and substantial computational resources. These constraints severely limit their applicability in real-world agricultural scenarios, particularly for deployment on low-power devices such as mobile phones, edge systems, and embedded platforms commonly used in smart farming applications.

To mitigate these challenges, recent research efforts have shifted toward ***lightweight CNN architectures***, including MobileNet and EfficientNet, which aim to reduce inference-time complexity through techniques such as depthwise separable convolutions and compound scaling. While these models successfully lower parameter count and computational overhead during inference, the majority of existing studies primarily emphasize model size reduction and runtime efficiency. Comparatively less focus has been placed on ***training efficiency, convergence speed, and optimization stability***, which are equally critical factors in practical agricultural environments where frequent retraining or domain adaptation may be required.

In real-world farming applications, rapid convergence and stable training behavior are essential, particularly when computational resources are limited or when models must be updated with new data. Motivated by these observations, this paper presents an ***efficient and lightweight convolutional neural network*** tailored specifically for potato leaf disease classification, with a strong emphasis on fast convergence and computational efficiency. The proposed architecture employs sequential convolutional blocks augmented with batch normalization and ReLU activation to ensure stable feature learning. Furthermore, global average pooling is adopted to replace parameter-heavy fully connected layers, significantly reducing model complexity. Adaptive optimization strategies, including learning rate scheduling and early stopping, are incorporated to accelerate convergence and prevent overfitting.

The effectiveness of the proposed approach is validated on a three-class potato leaf dataset consisting of Early Blight, Late Blight, and healthy leaf images. Experimental results demonstrate that the model achieves high classification accuracy within a limited number of training epochs while maintaining strong generalization capability. These findings underline the suitability of lightweight CNN architectures optimized for fast convergence as practical solutions for real-time potato disease diagnosis in resource-constrained agricultural settings.

TABLE I. Summary of Major Acronyms Used in This Paper

Acronym	Description
CNN	Convolutional Neural Network
GAP	Global Average Pooling
BN	Batch Normalization
ReLU	Rectified Linear Unit
Adam	Adaptive Moment Estimation Optimizer
LR	Learning Rate
Softmax	Probability Normalization Function
TP	True Positive
FP	False Positive
FN	False Negative
F1	F1-Score
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
GLCM	Gray-Level Co-occurrence Matrix
IoT	Internet of Things

The remainder of this paper is organized as follows. ***Section II*** presents a comprehensive review of related work on plant disease classification and lightweight deep learning approaches. ***Section III*** describes the proposed methodology, including dataset preparation, model architecture, and training strategy. ***Section IV*** reports the experimental results and performance evaluation. ***Section V*** provides a detailed discussion and comparative analysis of the findings. Finally, ***Section VI*** concludes the paper and outlines potential directions for future research.

In recent years, the problem of automated plant disease detection has attracted significant research interest, driven by the increasing availability of agricultural image datasets and rapid progress in computer vision technologies. Early investigations in this domain largely focused on conventional image processing techniques combined with classical machine learning algorithms to identify disease symptoms on plant leaves. These methods generally followed a multi-stage pipeline consisting of image preprocessing, segmentation, handcrafted feature extraction, and classification.

Initial studies employed preprocessing operations such as noise filtering, contrast enhancement, and color space transformation to improve image quality. Disease regions were commonly isolated using segmentation techniques including K-means clustering, Otsu's thresholding, and region-based methods. For feature representation, handcrafted descriptors such as texture features derived from the Gray-Level Co-occurrence Matrix (GLCM), color moments, histogram-based attributes, and shape descriptors were widely used. These features were subsequently classified using traditional machine learning models, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, and Random Forests. Although such approaches reported moderate classification accuracy—typically ranging between 80% and 90% under controlled experimental settings—their performance deteriorated in real-field conditions due to variations in illumination, background complexity, and leaf orientation. Furthermore, the reliance on manually engineered features limited their scalability and generalization capability.

A. Traditional Image Processing and Classical Machine Learning

Traditional plant disease detection frameworks were primarily based on explicit feature engineering pipelines. These approaches required careful selection of preprocessing steps and handcrafted features, making them sensitive to noise and environmental variations. While effective in laboratory-controlled environments, their dependence on domain-specific feature design restricted adaptability across different crops, disease types, and imaging conditions. Consequently, these limitations motivated the exploration of more robust, data-driven learning methods.

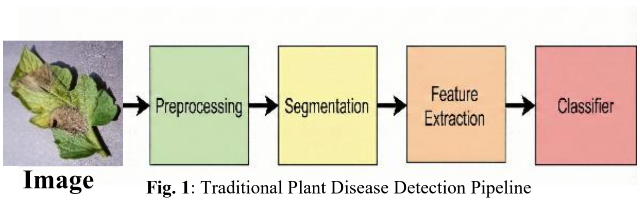

Fig. 1: Traditional Plant Disease Detection Pipeline

B. Deep Learning–Based Plant Disease Classification

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), introduced a paradigm shift in plant disease classification by enabling automatic learning of hierarchical feature representations directly from raw images. CNN-based approaches eliminated the need for handcrafted feature extraction and demonstrated substantial improvements over classical machine learning methods.

One of the earliest and most influential studies in this area was conducted by Mohanty *et al.*, who applied deep CNN architectures such as AlexNet and GoogLeNet to the PlantVillage dataset and reported classification accuracies exceeding 99%. However, their experimental analysis revealed a notable decline in performance when models trained on laboratory-controlled images were evaluated on real-field data, highlighting the impact of dataset bias and limited generalization. Similarly, Sladojevic *et al.* proposed a custom CNN architecture for multi-class plant disease recognition, demonstrating the feasibility of deep learning for agricultural applications. Despite achieving high accuracy, their model relied on deep convolutional stacks and required a large number of training epochs to reach convergence, leading to increased computational cost.

Subsequent studies followed similar trends, adopting deeper CNN architectures to enhance classification accuracy. While these models achieved strong performance, the improvements were often accompanied by higher computational complexity, longer training time, and increased memory requirements.

C. Transfer Learning and Deep Pre-trained Architectures

To overcome the challenge of limited labeled agricultural datasets, transfer learning emerged as a dominant strategy in plant disease classification. Numerous studies fine-tuned pre-trained deep models such as VGG16, ResNet50, Inception, and DenseNet using agricultural image datasets. By leveraging feature representations learned from large-scale datasets like ImageNet, these approaches consistently achieved classification accuracies in the range of 96% to 98%.

Despite their effectiveness, transfer learning–based architectures are characterized by large parameter counts and deep network hierarchies. For example, VGG16 contains approximately 138 million parameters, making it computationally expensive and unsuitable for deployment on low-resource agricultural devices. Although residual networks mitigate vanishing gradient issues through skip connections, they still demand substantial computational resources for both training and inference, thereby limiting their practicality for real-time agricultural applications.

D. Lightweight CNN Models and Efficiency-Oriented Approaches

learning–based models, recent research has increasingly focused on lightweight CNN architectures designed for efficient deployment on mobile and edge devices. Architectures such as MobileNet and EfficientNet employ design strategies including depthwise separable convolutions and compound scaling to reduce parameter count and inference latency.

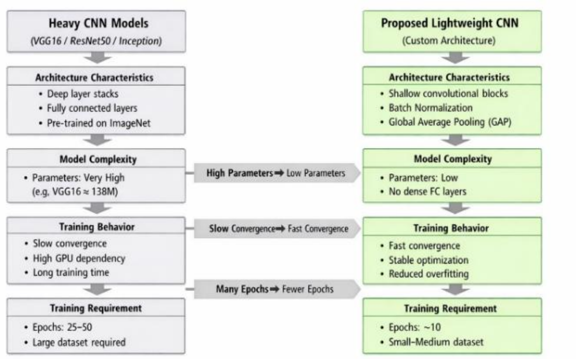

Fig. 2. Comparison between heavy CNN architectures and the proposed lightweight CNN in terms of parameters, convergence speed, and training efficiency

While these lightweight models offer improved efficiency compared to traditional deep CNNs, most existing studies primarily emphasize inference-time performance and final accuracy. Relatively limited attention has been given to training efficiency, convergence speed, and optimization stability. Moreover, many lightweight architectures incorporate complex structural components that may be unnecessary for low-cardinality classification tasks, such as three-class potato leaf disease detection. These factors indicate that existing solutions do not fully address the requirements of rapid training and efficient deployment in resource-constrained agricultural environments.

E. Summary and Research Gap

From the reviewed literature, it is evident that prior plant disease classification approaches either prioritize accuracy through deep and computationally expensive architectures or focus on inference efficiency without explicitly addressing training convergence and optimization behavior. There remains a clear research gap in developing task-specific CNN architectures that simultaneously balance classification accuracy, fast convergence, and computational efficiency.

Motivated by this gap, the present work proposes a custom lightweight convolutional neural network optimized for rapid convergence and reduced computational complexity. By integrating batch normalization to stabilize training and global average pooling to eliminate parameter-heavy fully connected layers, the proposed approach aims to bridge the gap between performance and efficiency in automated potato leaf disease classification.

III. METHODOLOGY

This section presents a detailed description of the dataset used for experimentation, the preprocessing pipeline, the proposed lightweight convolutional neural network architecture, the training strategy, and the mathematical formulation adopted to achieve rapid convergence and computational efficiency.

A. Dataset Description

The dataset employed in this study comprises labelled potato leaf images belonging to three distinct categories:

- ***Potato Early Blight***
- ***Potato Late Blight***
- ***Healthy Potato Leaves***

A total of ***2152 images*** were utilized for experimental evaluation. To ensure an unbiased assessment of model performance, the dataset was partitioned into training and validation subsets using an ***80:20 split***. All images were resized to a uniform spatial resolution of ***224 × 224 pixels***, ensuring compatibility with the proposed convolutional architecture and facilitating efficient batch-wise processing during training.

B. Data Preprocessing

Prior to training, all input images were normalized to standardize pixel intensity distributions, thereby improving numerical stability during optimization. To further enhance the generalization capability of the model and reduce the risk of overfitting, data augmentation strategies were applied exclusively during the training phase. These augmentation techniques included random horizontal flipping, random vertical flipping, and random rotation. By artificially increasing the diversity of training samples, the augmentation process improves robustness against variations in leaf orientation, scale, and appearance commonly observed in real-world agricultural images.

C. Proposed Lightweight CNN Architecture

The proposed model adopts a ***lightweight convolutional neural network architecture,*** specifically designed to achieve fast convergence while maintaining low computational complexity. The architecture consists of a sequence of convolutional blocks, each followed by batch normalization and ReLU activation to ensure stable and efficient feature learning. Spatial dimensionality reduction is performed using max-pooling layers, while *Global Average Pooling (GAP)* is employed to replace conventional fully connected layers, thereby significantly reducing the number of trainable parameters.

Architecture Details

- **Input:** RGB image of size $224 \times 224 \times 3$
- **Block 1:**
  Conv2D (32 filters, 3×3) → Batch Normalization → ReLU → Max Pooling (2×2)
  Output size: $112 \times 112 \times 32$
- **Block 2:**

Conv2D (64 filters, 3×3) → Batch Normalization → ReLU → Max Pooling (2×2)
Output size: $56 \times 56 \times 64$

- **Block3:**
  Conv2D (128 filters, 3×3) → Batch Normalization → ReLU
  Output size: $56 \times 56 \times 128$
- **Classifier:**
  Global Average Pooling → Fully Connected Layer → 3 output classes

The integration of GAP eliminates parameter-intensive dense layers, resulting in improved parameter efficiency while preserving discriminative spatial representations.

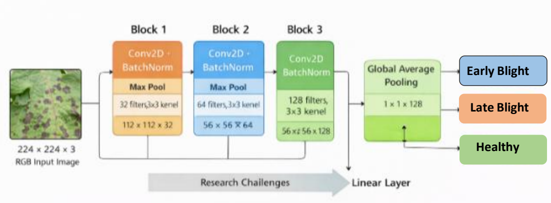

Fig. 3. Architecture of the proposed lightweight CNN for potato leaf disease classification

D. Training Strategy

The network parameters were optimized using the ***Adam optimizer***, which adaptively adjusts learning rates for individual parameters based on first- and second-order moment estimates. The ***categorical cross-entropy loss function*** was employed to optimize multi-class classification performance. To promote stable and rapid convergence, a ***ReduceLROnPlateau*** learning rate scheduler was used to dynamically decrease the learning rate when validation loss stagnated. Additionally, an ***early stopping mechanism*** was incorporated to terminate training once validation performance ceased to improve, thereby preventing overfitting and reducing unnecessary computational overhead.

TABLE II Training Hyperparameters of the Proposed CNN

Hyperparameter	Value
Input Image Size	(224 \times 224 \times 3)
Batch Size	32
Optimizer	Adam
Initial Learning Rate	0.001
Learning Rate Scheduler	ReduceLROnPlateau
LR Reduction Factor	0.1
Scheduler Patience	3 epochs
Loss Function	Categorical Cross-Entropy
Maximum Epochs	50
Early Stopping Patience	5 epochs
Weight Initialization	He Normal
Activation Function	ReLU
Number of Classes	3

E. Mathematical Framework

To address computational efficiency and convergence limitations reported in previous studies, the proposed CNN is formulated as a mapping function parameterized by a reduced parameter space Θ. The design incorporates two principal mechanisms: (i) mitigation of internal covariate shift through batch normalization, and (ii) parameter reduction through global average pooling

F. Convolutional Feature Extraction

Given an input image tensor X, the feature map $F_l$at layer $l$is computed using the convolution operation:

$$F_l = \sigma(W_l * X + b_l)$$

where $W_l$represents the learnable convolutional kernels, $b_l$denotes the bias term, $*$indicates the convolution operation, and $\sigma(\cdot)$is the Rectified Linear Unit (ReLU) activation function defined as:

$$\sigma(z) = \max(0, z)$$

The ReLU activation introduces non-linearity, promotes sparsity in feature representations, and alleviates the vanishing gradient problem during training.

G. Acceleration via Batch Normalization

To stabilize training dynamics and accelerate convergence, batch normalization (BN) layers are applied after each convolutional layer. Given a mini-batch $B = \{x_1, x_2, \dots, x_m\}$, batch normalization is defined as:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$y_i = \gamma \hat{x}_i + \beta$$

where $\mu_B$and σ_B^2represent the batch mean and variance, respectively, $\gamma$and $\beta$are learnable scaling and shifting parameters, and $\epsilon$is a small constant introduced for numerical stability. This normalization constrains feature distributions to a predictable range, enabling the use of higher learning rates and reducing the number of epochs required to reach convergence.

H. Efficiency via Global Average Pooling

In conventional CNN architectures, fully connected layers account for a substantial portion of trainable parameters. To overcome this limitation, the proposed model employs ***Global Average Pooling (GAP).*** For a feature map $A \in \mathbb{R}^{h \times w \times d}$, GAP computes a vector $v \in \mathbb{R}^d$as:

$$v_k = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} A_{i,j,k}$$

This operation reduces parameter complexity from $O(M \cdot N_{dense})$to $O(d \cdot C)$, where $M = h \cdot w \cdot d$and Cdenotes the number of output classes.

I. Optimization Objective

The classification problem is modeled using the ***Softmax*** function. For an output vector z, the probability of assigning an input image X to class j is given by:

$$P(y = j \mid X) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

where K denotes the total number of disease classes.

IV. EXPERIMENTAL

This section reports the experimental evaluation of the proposed lightweight convolutional neural network, focusing on training behavior, convergence characteristics, and classification performance on the validation dataset.

A. Training and Validation Performance

The proposed lightweight CNN demonstrates ***rapid and stable convergence***, achieving optimal performance within approximately ***10 training epochs***. The training and validation loss trajectories exhibit a smooth and monotonic decrease across epochs, indicating well-behaved learning dynamics and stable optimization throughout the training process.

A key observation is that the validation loss consistently follows the training loss with only a minimal gap, and no noticeable divergence is observed during training. This behavior indicates that the proposed model does not suffer from overfitting and is capable of generalizing effectively to unseen data. The integration of batch normalization layers, global average pooling, and early stopping contributes significantly to stabilizing the optimization process and preventing excessive memorization of training samples.

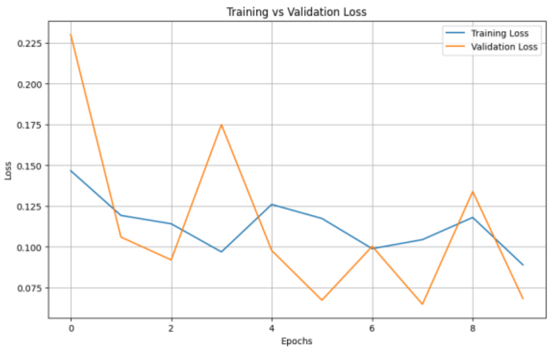

Fig. 4. Training and validation loss curves of the proposed lightweight CNN.

In addition, the validation accuracy curve closely tracks the training accuracy trend across epochs, further confirming the model's ability to learn discriminative feature representations without compromising generalization performance. The ability to reach convergence within a limited number of epochs highlights the ***computational efficiency*** of the proposed architecture, making it suitable for deployment in ***low-resource and real-time agricultural environments***.

B. Classification Performance

The classification capability of the proposed model was assessed on the validation dataset using standard evaluation metrics, including ***precision, recall, F1-score, and overall accuracy***. The experimental results show that the model attains a ***high overall classification accuracy of 98%***, with consistently strong precision and recall values across all disease categories.

Table III. The detailed classification report is summarized in

Class	Precision Score	Recall	F1-Score
Potato Early Blight	0.99	0.98	0.98
Potato Late Blight	0.96	0.99	0.97
Healthy Potato	1	0.9	0.95
Overall Accuracy			**0.98**

In particular, the ***Early Blight*** and ***Late Blight*** classes achieve high recall scores, indicating that the model is highly effective in identifying diseased leaves. This property is especially important for practical agricultural disease diagnosis, where failure to detect infected plants can lead to rapid disease spread and significant crop losses.

A comparatively lower recall is observed for the ***Healthy Potato*** class, which can be attributed to the visual resemblance between healthy leaves and those exhibiting mild disease symptoms. Despite this, the precision for the healthy class remains high, suggesting that predictions labeled as healthy are highly reliable. Overall, the balanced performance across classes reflects the robustness of the proposed lightweight CNN.

C. Confusion Matrix Analysis

Further insight into the classification behavior is provided through confusion matrix analysis. The confusion matrix of the proposed lightweight CNN is illustrated in **Fig. 5**. The results indicate that the majority of validation samples are correctly classified across all three classes.

Only a small number of misclassifications are observed, primarily between ***Early Blight*** and ***Late Blight***, which share visually similar disease patterns. Importantly, confusion between diseased and healthy leaves remains minimal, demonstrating the model's ability to effectively separate healthy samples from infected ones. This behavior confirms that the proposed CNN learns robust and discriminative spatial features relevant to potato leaf disease identification, supporting its reliability for real-world agricultural deployment.

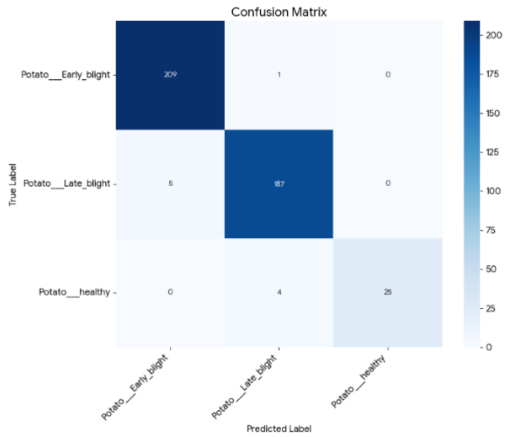

Fig. 5. Confusion matrix of the proposed lightweight CNN on the validation set

D. Effect of Lightweight Design Choices

The fast convergence and stable training behaviour observed in the experiments can be largely attributed to the architectural design choices adopted in the proposed model. Batch normalization improves gradient stability and reduces internal covariate shift, allowing the network to converge more quickly under higher learning rates. In addition, replacing conventional fully connected layers with global average pooling substantially reduces the number of trainable parameters, which helps mitigate overfitting and enhances generalization, particularly when training on limited agricultural datasets.

Together, these design decisions enable the proposed lightweight CNN to achieve high classification performance with reduced computational cost, validating the effectiveness of the architecture for efficient plant disease classification.

V. DISCUSSION

The experimental findings clearly indicate that achieving high classification accuracy does not necessarily require deep or computationally intensive neural network architectures. The proposed lightweight CNN demonstrates that carefully designed architectural choices can effectively balance performance and efficiency. In particular, the integration of global average pooling significantly reduces model complexity by eliminating parameter-heavy fully connected layers, while adaptive training strategies enable rapid convergence without compromising classification accuracy.

The strong generalization performance observed across validation experiments suggests that the proposed model successfully learns discriminative spatial features relevant to potato leaf disease patterns. Batch normalization plays a crucial role in stabilizing gradient flow and accelerating optimization, whereas early stopping prevents overfitting by limiting unnecessary training iterations. Together, these design choices allow the model to converge within a limited number of epochs while maintaining robust performance across disease categories.

From a practical perspective, the ability to achieve reliable disease classification with reduced training time and lower computational cost makes the proposed approach particularly suitable for real-time agricultural applications. Compared to deep and transfer learning–based models, the proposed lightweight CNN offers a favourable trade-off between accuracy, convergence speed, and resource efficiency, which is essential for deployment in low-resource and edge-based farming environments.

VI. LIMITATIONS AND FUTURE WORK

Although the proposed approach demonstrates promising performance, certain limitations remain. The dataset used in this study primarily consists of images captured under controlled conditions, with limited variation in background, lighting, and environmental factors. As a result, the model's robustness under complex real-field conditions has not been fully explored.

Future research will focus on addressing these limitations by evaluating the proposed framework on real-world agricultural datasets containing diverse illumination conditions, cluttered backgrounds, and varying leaf orientations. In addition, extending the current framework to support ***multi-crop and multi-disease classification*** represents a natural progression of this work. Further investigation will also explore ***deployment on edge devices and mobile platforms***, enabling real-time disease diagnosis directly in the field. These extensions aim to enhance the practical applicability and scalability of the proposed system within smart agriculture ecosystems.

VII. CONCLUSION

This paper presented an efficient convolutional neural network–based approach for potato leaf disease classification, with a specific focus on fast convergence and computational efficiency. By combining lightweight architectural design with adaptive optimization strategies, the proposed model achieved a validation accuracy of ***98% within only 10 training epochs***, demonstrating its effectiveness for automated disease diagnosis.

The experimental results confirm that compact CNN architectures, when carefully designed, can deliver high classification performance while significantly reducing training time and computational overhead. These findings highlight the potential of lightweight deep learning models as practical and scalable solutions for crop disease detection in resource-constrained agricultural environments. The proposed approach therefore represents a promising step toward the development of real-time, intelligent plant health monitoring systems for precision agriculture.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, pp. 1–10, 2016.

[2] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based

recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3289801, 2016.
[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. Int. Conf. on Learning Representations (ICLR)*, 2015.
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
[5] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[6] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proc. Int. Conf. on Machine Learning (ICML)*, pp. 6105–6114, 2019.
[7] T.-Y. Lin *et al.*, "Feature pyramid networks for object detection," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2117–2125, 2017.
[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proc. Int. Conf. on Learning Representations (ICLR)*, 2015