International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 02 | Feb - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Comparative Study of a Custom CNN and Pre-Trained Deep Learning Models for Potato Leaf Disease Classification

Hima Gandhi[1], Madan Lal[2], Kanwal Preet Singh Attwal[3]

[1]Department of Computer Science & Engineering, Punjabi University, Patiala, Punjab, India
[2,3]Department of Computer Science & Engineering, Punjabi University, Patiala, Punjab, India

---

Abstract - Accurate detection of potato leaf diseases is critical for minimizing yield losses and supporting sustainable agriculture. This study evaluates the performance of multiple deep learning architectures for classifying three categories of potato leaves: Early Blight, Late Blight, and Healthy. Six state-of-the-art models, including VGG19, DenseNet121, InceptionV3, MobileNetV3Large, Vision Transformer (ViT), and a Custom CNN trained from scratch were assessed using metrics such as Accuracy, Precision, Recall, and F1-score. A two-stage training strategy was employed, leveraging pre-trained ImageNet weights for transfer learning followed by fine-tuning of higher layers to adapt features to domain-specific patterns. Experimental results demonstrated that VGG19 achieved the highest overall performance, with 97.83% accuracy, 0.9419 precision, 0.9721 recall, and 0.9559 F1-score. DenseNet121 and InceptionV3 also exhibited strong performance, while lightweight and transformer-based models (MobileNetV3Large and ViT) showed moderate results due to the limited dataset size. The Custom CNN achieved high validation accuracy but slightly lower test performance, highlighting the advantage of transfer learning for improved generalization. These findings confirm that deep convolutional architectures combined with transfer learning provide robust and reliable solutions for automated potato leaf disease classification.

***Key Words*: Potato leaf disease detection, PlantVillage Dataset, Transfer Learning, Fine-Tuning, CNN**

1. INTRODUCTION

1.1 Background and Motivation

The agriculture sector constitutes a cornerstone of India's economic development, significantly influencing economic growth, ensuring food security, increasing employment and fostering rural development. Potatoes (Solanum tuberosum), belonging to the Solanaceae family, are a major staple crop that contributes significantly to global nutrition and agricultural income. They are consumed worldwide due to their high carbohydrate, vitamin, and mineral content. However, potato production is highly vulnerable to diseases that cause severe yield losses. Diseases such as late blight, early blight, bacterial wilt, viral infections, nematode infestations, and pest damage often devastate potato crops, leading to economic losses for farmers and threatening food security. Main diseases of potato plants are Phytophthora infestans (Late Blight) and Alternaria solani (Early Blight). Moreover, many farmers, especially in rural areas, lack access to trained pathologists, resulting in delayed or incorrect diagnosis. Recent advances in machine learning and computer vision have shown promise in automating disease detection from plant leaf images. In particular, Convolutional Neural Networks (CNNs) and their modern variants have become the foundation of automated plant disease detection due to their ability to learn hierarchical and discriminative features from images. Pre-trained architectures such as VGG19, DenseNet121, InceptionV3, MobileNetV3Large, and Vision Transformers (ViT) leverage knowledge from large-scale datasets such as ImageNet, enabling effective feature extraction even with limited leaf images, while custom CNNs can be specifically designed to capture subtle morphological variations unique to potato leaf diseases. Fine-tuning these pre-trained models allows the network to adapt to domain-specific characteristics, improving classification performance across multiple disease categories. By combining the strengths of pre-trained networks with specialized custom architectures, the system achieves robust detection, resilience to visual noise, and higher accuracy, making it suitable for real-time deployment in agricultural settings. This approach provides a scalable and efficient solution for automated disease monitoring, facilitating early diagnosis and timely intervention, which are critical for minimizing crop losses and supporting sustainable potato cultivation. Motivated by these advances, this work presents a comprehensive potato leaf disease classification system that leverages state-of-the-art deep learning techniques, enabling rapid, accurate, and scalable disease detection, thereby supporting sustainable agricultural practices and improving overall crop productivity.

1.2 Importance of Potato Leaf Disease Classification

Potato is one of the most widely cultivated and consumed food crops in the world, playing a vital role in global food security and agricultural economies. Countries such as India, China, and Russia are among the leading producers of potatoes, where the crop serves as a staple food and a major source of income for farmers. However, potato plants are highly susceptible to several leaf diseases, particularly early blight and late blight, which can significantly reduce yield and crop quality if not detected and managed in time. One of the most devastating potato diseases is late blight, historically associated with the Great Irish Famine [1] , which led to massive crop failures and food shortages. Even today, such diseases continue to pose a serious threat to agricultural productivity, especially in regions with favorable conditions for fungal growth. Early detection of leaf diseases enables timely application of control measures such as fungicides, proper irrigation, and crop management practices, thereby minimizing crop loss. Manual inspection of potato fields is slow, labor-intensive, and often inaccurate, particularly for large farms or early disease stages. Farmers may also misidentify diseases, leading to incorrect treatments and crop losses. Automated detection using image processing and deep learning provides a fast and accurate alternative by analyzing leaf images and classifying diseases in real time. This enables early intervention, reduces excessive pesticide use, lowers costs, and supports sustainable and precision farming practices. Therefore, this study explores the use of deep learning–based image classification techniques to develop an accurate and automated system for potato leaf disease detection.

1.3 Overview of Implementation

This research investigates transfer learning approaches for detecting and classifying three categories of potato leaf conditions: Early Blight, Late Blight, and Healthy. The contributions of this study can be summarized as follows. First, a well-structured dataset of potato leaf images was prepared and systematically divided into training, validation, and testing sets in a 70-15-15 ratio to ensure balanced and unbiased evaluation. Second, several state-of-the-art Convolutional Neural Network (CNN) architectures, including VGG19, InceptionV3, DenseNet121 and MobileNetV3Large, along with a Vision Transformer (ViT), were implemented using transfer learning to leverage pre-trained ImageNet knowledge for effective feature extraction. Third, the models were trained in a two-stage strategy consisting of transfer learning followed by fine-tuning. In the first stage, the convolutional or transformer backbone was frozen and only the newly added classification layers were trained. In the second stage, the top layers or transformer blocks were unfrozen and fine-tuned using a lower learning rate, allowing the models to adapt high-level features specifically to potato leaf disease patterns and improve generalization performance. Fourth, the performance of these models was thoroughly evaluated using multiple metrics such as accuracy, precision, recall, F1-score, confusion matrices, and training curves to provide a comprehensive analysis of classification capability. Finally, among the tested models, VGG19 achieved the best overall performance, demonstrating superior generalization and reliability for multi-class classification of potato leaf diseases, indicating its potential for practical deployment in agricultural decision-support systems.

2. Related Work

Various techniques have been applied in agriculture for the detection of plant diseases and pests. These include deep learning and image processing methods, as well as traditional machine learning approaches that have been widely used in this domain. Too et al. [2] conducted a comparative study on fine-tuning deep learning architectures for plant disease identification. Their work focused on developing an automatic and accurate image-based classification system using state-of-the-art convolutional neural networks. The study evaluated multiple architectures, including VGG16, InceptionV4, ResNet (50, 101, and 152 layers), and DenseNet121, using the PlantVillage dataset containing 38 classes of healthy and diseased leaves from 14 plant species. The authors emphasized the importance of fast and reliable disease detection systems to support early intervention and improve food security. Experimental results showed that DenseNet121 consistently improved accuracy as training progressed, without significant overfitting or performance degradation. In addition, it required fewer parameters and reasonable computational time compared to other models. DenseNet121 achieved the best performance, with a test accuracy of 99.75%, outperforming all other evaluated architectures. The models were implemented and trained using the Keras framework with the Theano backend. Sauda et al. [3] conducted a comparative study using transfer learning for tomato leaf disease classification. The authors utilized three pre-trained deep learning architectures: Inception-v3, ResNet-50, and Inception-ResNet-v2 originally trained on the ImageNet dataset. The final classification layers of these models were replaced to accommodate four target classes (three disease categories and one healthy class). The models were trained for 100 epochs with a mini-batch size of 32 using images from the

PlantVillage dataset. Experimental results showed that all three architectures achieved high performance, with average training accuracies exceeding 99%. Inception-v3 achieved the highest performance, obtaining an average training accuracy of 99.5% and validation accuracy of 98.3%. ResNet-50 followed closely with 99.4% training and 98.1% validation accuracy. Inception-ResNet-v2 achieved 99.1% training and 97.5% validation accuracy, demonstrating gradual performance improvement across epochs. The study concluded that transfer learning with deep convolutional neural networks is highly effective for plant disease detection, with Inception-v3 outperforming the other evaluated models in terms of overall accuracy. Tambe et al. [4] proposed a Convolutional Neural Network (CNN) based approach for potato leaf disease classification. The study focused on detecting Early Blight, Late Blight, and healthy leaves using image data. The authors implemented preprocessing steps, including resizing, normalization, and augmentation, to improve model generalization. A custom CNN model was then trained on the pre-processed dataset, and its performance was evaluated on a separate test set. Experimental results demonstrated that the CNN achieved an overall accuracy of 99.1%, effectively classifying all three categories even in cases of severe disease infections. The study highlighted that deep learning models, specifically CNNs, can automatically extract relevant features from leaf images, eliminating the need for handcrafted feature engineering. The authors concluded that CNN-based approaches provide a reliable and efficient solution for automated potato disease detection, which is crucial for minimizing yield losses and supporting precision agriculture practices. Sohel et al. [5] investigated the application of deep learning models for potato pest classification. The study aimed to detect pests affecting eight prevalent potato species using image data collected from multiple sources. The authors employed several image pre-processing techniques, including resizing, normalization, and enhancement, to improve image quality and ensure compatibility with deep learning models. Three convolutional neural network (CNN) architectures: InceptionV3, VGG-16, and MobileNetV2 were evaluated for their classification performance. Experimental results demonstrated that VGG-16 achieved the highest accuracy of 94.44%, outperforming MobileNetV2, which reached 75%, and InceptionV3, which attained 58%. The study emphasized that preprocessing plays a crucial role in enhancing model performance by reducing noise and improving feature representation. Shabrina et al. [6] introduced a novel dataset tailored for potato leaf disease detection under real-world conditions, addressing a key limitation in existing research that largely depends on controlled datasets such as PlantVillage. The authors argued that images captured in controlled environments, clean backgrounds and fixed lighting do not adequately represent the variabilities encountered in actual field settings. To overcome this, they compiled a dataset of 3,076 high-resolution leaf images collected from multiple potato farms in Central Java, Indonesia, under uncontrolled environmental conditions with diverse backgrounds, angles, and lighting variations. The dataset includes seven distinct classes: virus, phytophthora, nematode, fungi, bacteria, pest, and healthy-which provides broader disease diversity compared to traditional datasets that typically include only a few fungal disease categories. This richer dataset facilitates more realistic evaluations of machine learning and deep learning models for potato disease identification. The work emphasizes that using images from uncontrolled environments can help researchers develop and benchmark more robust classification models capable of handling background clutter, occlusion, and other natural variabilities encountered in practical agricultural scenarios. Thus, this dataset serves as a valuable resource for advancing automatic potato leaf disease classification systems and supports the development of more effective precision agriculture tools. Mhala et al. [7] proposed a deep learning-based approach for the detection and classification of six major potato leaf diseases, including bacteria, viruses, fungi, phytophthora, pests, and nematodes. The study addressed challenges associated with class imbalance by applying strategic data augmentation, L2 regularization, and transfer learning. Three pre-trained convolutional neural networks: DenseNet201, ResNet152V2, and NasNetMobile were fine-tuned on a dataset of 3,076 images collected under real-world, uncontrolled conditions. DenseNet201 achieved the highest baseline accuracy of 77.14% and further improved to an average accuracy of 81.31% with data augmentation and k-fold cross-validation, outperforming previously reported results by 7.68%. In contrast, NasNetMobile and ResNet152V2 experienced performance declines due to overfitting and their limited capacity to handle the increased variability introduced by augmentation. The study emphasizes the importance of addressing class imbalance and leveraging appropriate regularization and augmentation techniques to enhance the reliability and robustness of disease detection models. These findings demonstrate the potential of deep learning frameworks, particularly DenseNet201, in providing scalable and effective solutions for automated potato disease monitoring, thereby supporting sustainable agricultural practices and minimizing crop losses. Chowdhury and Das [8] investigated the application of deep learning models for potato leaf disease detection, focusing on Early Blight, Late Blight, and healthy leaves. The study employed a dataset of 3,251 leaf images, which were pre-processed by resizing to 224 × 224 pixels and normalizing pixel values to enhance compatibility with deep learning models. Two architectures: a standard Convolutional Neural Network (CNN)

and ResNet50 were implemented and trained for 20 epochs each. Experimental results demonstrated that ResNet50 outperformed the CNN model, achieving a validation accuracy of 97%, whereas the CNN model reached only 76%. The authors also provided detailed analyses of classification reports, confusion matrices, and graphical representations of training and validation loss, highlighting the superior learning capability and stability of ResNet50 in tracking validation performance. This study reinforces the efficacy of transfer learning and deep convolutional architectures for automated potato leaf disease detection, providing a reliable framework for precision agriculture applications and early disease management strategies.

3. Materials and Methods

The potato leaf disease classification system proposed follows a structured pipeline. The input leaf images are processed through dataset splitting, preprocessing, augmentation, feature extraction via pre-trained and custom CNN models, and classification into disease categories. This section describes each component of the pipeline, with an overview of the process presented in Figure 1.

The flowchart consists of the following steps:

Start: Input potato leaf dataset (2,152 images, 3classes).

Step 1: Split dataset → 70% training, 15% validation, 15% testing.

Step 2: Preprocess images → Rescale, Rotate, Shift, Shear, Zoom, Flip

Step 3: Load pre-trained CNN architectures (VGG19, DenseNet121, InceptionV3, MobileNetV3Large, Vision Transformer (ViT)), along with a custom CNN model.

Step 4: Add custom classification layers (Global Average Pooling, Dense, Dropout, SoftMax).

Step 5: Train models for pre-trained CNN Models (Stage 1: freeze base, Stage 2: fine-tune 20% of top layers).

Step 6: Monitor training with callbacks (EarlyStopping, ModelCheckpoint, ReduceLROnPlateau).

Step 7: Evaluate models using Accuracy, Precision, Recall, F1-score, Confusion Matrix.

End: Identify best-performing model (VGG19).

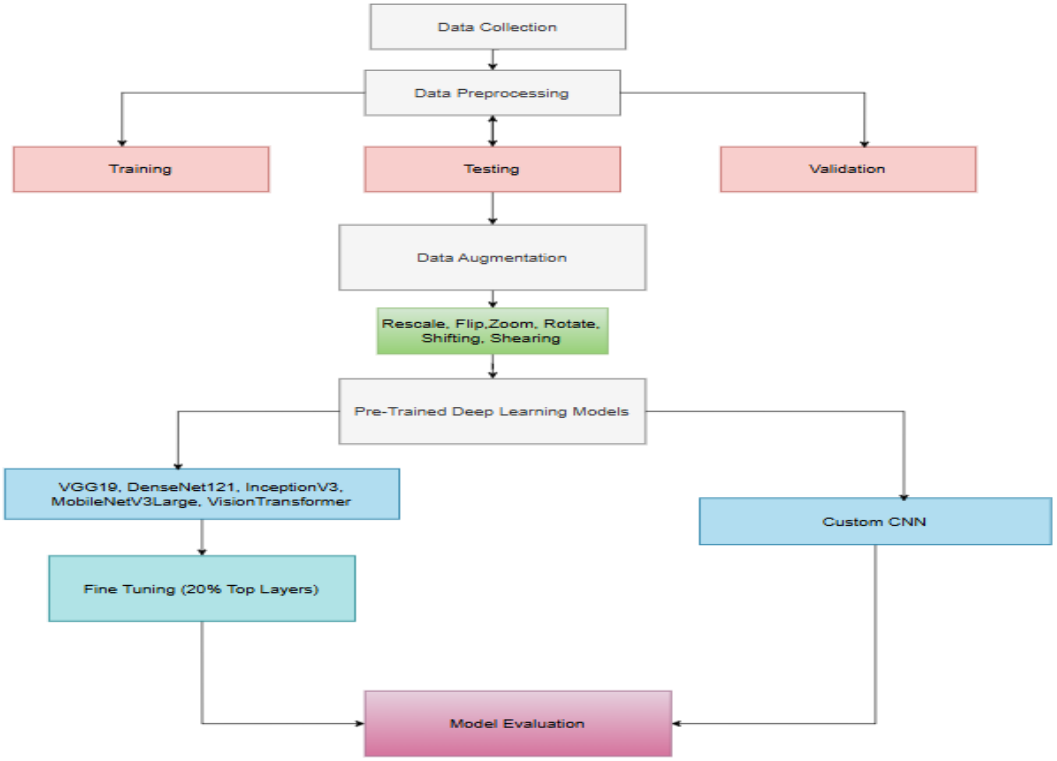

Figure 1: System Flowchart

3.1 Dataset

3.1.1 Description of PlantVillage Dataset

Deep learning models were trained and evaluated on potato leaf images with the objective to accurately classify and identify diseases on test samples. For the present study, an openly accessible dataset from PlantVillage was utilized. The

dataset used consists of 2,152 sample of potato leaf images with 3 categories of Early Blight, Late Blight and Healthy samples. Figure 2 shows a sample of dataset used [9].

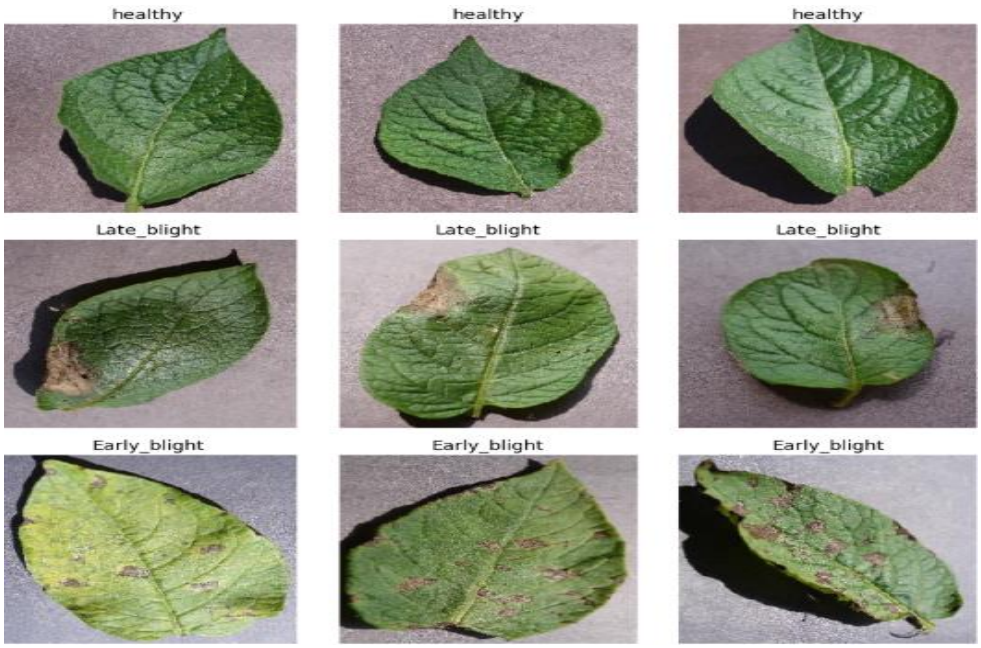

Figure 2: Sample of Potato Leaf Dataset

3.1.2 Data Splitting

The dataset was stratified into train, test and validation sets with the ratio of 70-15-15. This way each set receives equal representation of classes and, therefore, the model can be judged fairly while being tested. Approximately, 15% of the images or about 323 make it to test and validation set, and about 1,506 images, stay with the train set. The test set is used for prediction and evaluation of the models. By keeping it stratified Early Blight, Late Blight and Healthy label distribution intact, this acts as an advantage for the model to generalize across the dataset.

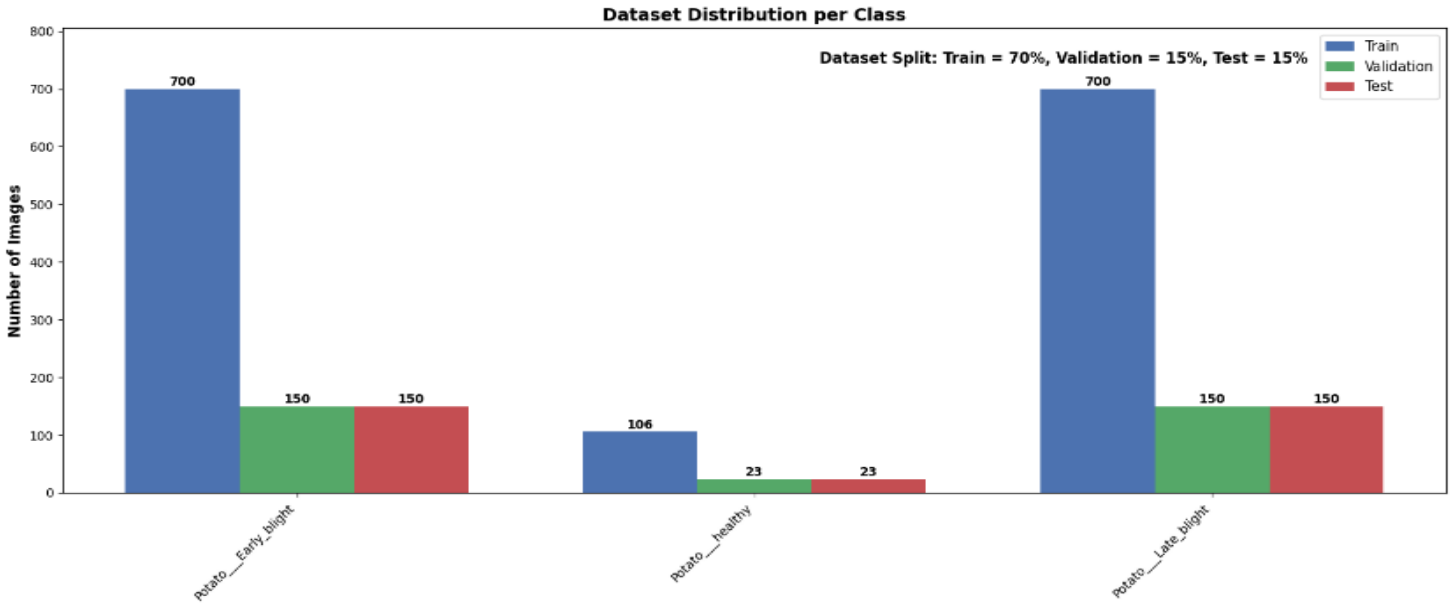

Figure 3: Data Splitting

3.2 Data Augmentation

To improve the generalization capability of the models and reduce overfitting, data augmentation techniques were applied to the training dataset. All images were first resized to 224×224 pixels to ensure uniform input dimensions across

the deep learning architectures. Image transformations were then performed using a combination of geometric and photometric modifications, including random rotations of up to 30°, horizontal and vertical shifts up to 20%, shearing, zooming, and horizontal flipping. The images were rescaled to normalize pixel values to the range [0,1], and empty pixels resulting from transformations were filled using the nearest pixel approach. The specific augmentation parameters and their configurations are summarized in Table I. These augmentation strategies artificially expanded the training dataset and exposed the models to diverse variations of leaf images, enabling more robust feature learning and improving the accuracy of disease classification on unseen samples.

Table no 1: Summary of Data Augmentation Techniques and Parameters

Augmentation Type	Parameters / Range	Description
Rescaling	1./255	Normalize pixel values to [0,1]
Rotation	0–30°	Randomly rotate images
Width Shift	0–20%	Horizontal translation
Height Shift	0–20%	Vertical translation
Shear	0–20%	Shear transformation
Zoom	0–20%	Random zoom in/out
Horizontal Flip	True	Random horizontal flipping
Fill Mode	nearest	Fill empty pixels after transformation

3.3 State-of-the-art Deep Learning Image Classifiers

3.3.1 VGG19 Model

VGG19 architecture is a deep convolutional neural network consisting of 19 weight layers, characterized by sequential 3×3 convolutional filters and max-pooling layers. In this work, the pre-trained VGG19 model was loaded without the top fully connected layers (include_top=False).

All convolutional layers were initially frozen during Stage-1 transfer learning to retain ImageNet-learned features. A custom classification head was added, consisting of:

- Global Average Pooling layer
- Dense layer (512 neurons, ReLU activation)
- Dropout layer (0.5)
- Final Dense layer with SoftMax activation

The model was trained using the categorical cross-entropy loss function. During Stage-2 fine-tuning, the top 20% of convolutional layers were unfrozen to adapt high-level features to potato leaf disease characteristics. VGG19 demonstrated strong generalization performance and achieved one of the highest test accuracies among all evaluated models.

3.3.2 DenseNet121

The DenseNet121 architecture introduces dense connectivity, where each layer receives feature maps from all preceding layers. This design improves feature reuse and gradient flow. The pre-trained DenseNet121 model was loaded without its top layers. All dense blocks were frozen during transfer learning. A custom classification head was added consisting of Global Average Pooling, Dense (512 units), Dropout (0.5), and a SoftMax layer. The model was trained using the categorical cross-entropy loss function. During fine-tuning, the top 20% of dense layers were unfrozen. DenseNet121 achieved high validation and test accuracy with efficient parameter utilization, demonstrating stable convergence behavior.

3.3.3 InceptionV3

The InceptionV3 model employs inception modules that capture multi-scale spatial information using parallel convolutional filters. The pre-trained InceptionV3 backbone was used without the top classification layers. The base model was initially frozen. A deeper classification head was designed consisting of:

- Global Average Pooling
- Dense layer (1024 units, ReLU)
- Dropout (0.3)
- Dense layer (512 units, ReLU)
- Dropout (0.3)
- SoftMax output layer

The model was trained using the categorical cross-entropy loss function. During fine-tuning, upper inception modules were unfrozen. InceptionV3 showed strong performance and stable learning behavior across epochs.

3.3.4 MobileNetV3Large

The MobileNetV3Large architecture is designed for lightweight and efficient computation using depthwise separable convolutions and squeeze-and-excitation modules. The pre-trained MobileNetV3Large backbone was used without the classification head. A custom classification head consisting of Global Average Pooling, Dense (512 units), Dense (256 units), Dropout layers (0.3), and a SoftMax layer was appended. The model was trained using the categorical cross-entropy loss function. Although computationally efficient, MobileNetV3Large achieved moderate classification performance compared to deeper architectures.

3.3.5 Vision Transformer

The Vision Transformer is a transformer-based architecture that applies self-attention mechanisms to image patches. In this study, a pre-trained ViT base model (vit-base-patch16-224) was utilized. During Stage-1 transfer learning, all transformer encoder blocks were frozen, and a classification head consisting of:

- Dense layer (512 units, ReLU)
- Dropout (0.4)
- SoftMax output layer

was added.

The model was trained using the categorical cross-entropy loss function. In Stage-2 fine-tuning, the top 20% of transformer blocks were unfrozen. Fine-tuning improved validation accuracy, demonstrating the importance of adapting high-level attention features for plant disease classification.

3.3.6 Custom CNN

A custom convolutional neural network was also designed to evaluate performance without pre-trained weights. The architecture consists of multiple convolutional blocks with increasing filter sizes (32, 64, 128, 256, 512), each followed by max-pooling layers. Dropout regularization was applied to reduce overfitting. The fully connected section included:

- Flatten layer
- Dense layer (1500 units, ReLU)
- Dropout (0.4)
- SoftMax output layer

The model was trained from scratch using the categorical cross-entropy loss function and Adam optimizer as shown in Figure 4 :

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d_188 (Conv2D)	(None, 224, 224, 32)	896
conv2d_189 (Conv2D)	(None, 222, 222, 32)	9,248
max_pooling2d_8 (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_190 (Conv2D)	(None, 111, 111, 64)	18,496
conv2d_191 (Conv2D)	(None, 109, 109, 64)	36,928
max_pooling2d_9 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_192 (Conv2D)	(None, 54, 54, 128)	73,856
conv2d_193 (Conv2D)	(None, 52, 52, 128)	147,584
max_pooling2d_10 (MaxPooling2D)	(None, 26, 26, 128)	0
conv2d_194 (Conv2D)	(None, 26, 26, 256)	295,168
conv2d_195 (Conv2D)	(None, 24, 24, 256)	590,080
max_pooling2d_11 (MaxPooling2D)	(None, 12, 12, 256)	0
conv2d_196 (Conv2D)	(None, 12, 12, 512)	1,180,160
conv2d_197 (Conv2D)	(None, 10, 10, 512)	2,359,808
max_pooling2d_12 (MaxPooling2D)	(None, 5, 5, 512)	0
dropout_9 (Dropout)	(None, 5, 5, 512)	0
flatten (Flatten)	(None, 12800)	0
dense_15 (Dense)	(None, 1500)	19,201,500
dropout_10 (Dropout)	(None, 1500)	0
dense_16 (Dense)	(None, 3)	4,503

Total params: 71,754,683 (273.72 MB)
Trainable params: 23,918,227 (91.24 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 47,836,456 (182.48 MB)

Figure 4: Layer-wise architecture of the proposed Custom CNN model

3.4 Model Training

3.4.1 Model Loss Function and Optimizer

The models were trained using the categorical cross-entropy loss function, which is suitable for multi-class classification problems. This loss function measures the difference between the predicted probability distribution and the true class labels, guiding the model toward improved classification accuracy. The Adam optimizer was employed due to its adaptive learning rate capabilities and efficient convergence. An initial learning rate of 1e-4 was used during the initial training stage to enable stable and efficient optimization. During the fine-tuning stage, the learning rate was reduced to 1e-5 to allow more precise weight updates and prevent disruption of previously learned features. Training was conducted using mini-batches of 32 images, which provided a balance between computational efficiency and stable gradient updates.

3.4.2 Callbacks and Early Stopping

To enhance training efficiency and prevent overfitting, several callback mechanisms were employed:

ModelCheckpoint: This callback saved the model weights corresponding to the highest validation accuracy, ensuring that the best-performing model was preserved.

ReduceLROnPlateau: This callback monitored the validation loss and reduced the learning rate by a factor of 0.2 if no improvement was observed for 3 consecutive epochs, with a minimum learning rate of 1e-6.

EarlyStopping: Training was halted if the validation loss did not improve for 5 consecutive epochs, and the model weights from the best epoch were restored automatically.

Although a maximum of 100 epochs was allowed during training, the EarlyStopping mechanism typically terminated the process earlier, reducing computational cost while maintaining optimal model performance.

3.4.3 Fine-tuning the models

Fine-tuning is a concept of transfer learning. Transfer learning is a machine learning technique in which knowledge gained while solving one problem is applied to a related task or domain. In deep learning, the initial layers of a network learn general features such as edges, textures, and shapes, while the deeper layers capture more task-specific patterns. During transfer learning, the original classification layers of a pre-trained network are replaced with new layers suitable for the target task, and the network is retrained on the new dataset. Fine-tuning allows the model to adapt learned features to the specific problem while requiring significantly less training time and data compared to training from scratch. In

this study, several state-of-the-art deep learning models pre-trained on the ImageNet dataset were fine-tuned for potato leaf disease classification. The ImageNet dataset contains over one million images across 1000 classes, enabling the models to learn rich and generalized visual features. The target dataset used in this research was significantly smaller; therefore, transfer learning was adopted to improve training efficiency and classification performance. The pre-trained models used in this study included VGG19, DenseNet121, InceptionV3, MobileNetV3Large, and Vision Transformer (ViT). For each model, the original top classification layers were removed and replaced with new fully connected layers followed by a SoftMax output layer corresponding to the number of disease classes. During the initial training phase, the convolutional base of each pre-trained model was frozen, and only the newly added classification layers were trained. This allowed the models to preserve the general feature representations learned from ImageNet while adapting to the potato leaf disease dataset. After this stage, the upper layers of the networks were unfrozen, and the models were fine-tuned using a lower learning rate of 1e-5 to enable more precise weight updates without disrupting the previously learned features. All models were trained using the Adam optimizer, with an initial learning rate of 1e-4 during the feature-extraction phase. Data augmentation was applied during training to improve generalization and reduce overfitting.

Table no 2: Transfer Learning Phase – Accuracy and Loss for Training and Validation

Model And Specifications					**Transfer Learning Phase**			
Models	**Convolutional Layers**	**Total Layers**	**Frozen Layers**	**Parameters (Millions)**	**Training Accuracy %**	**Validation Accuracy%**	**Training Loss**	**Validation Loss**
VGG19	16	22	17	20.02M	91.77	91.64	0.2066	0.2136
DenseNet121	120	427	341	7.04M	98.54	96.59	0.0512	0.0912
InceptionV3	94	311	248	21.80M	93.63	94.74	0.1704	0.1583
MobileNetV3Large	62	187	149	3.00M	75.03	69.04	0.6467	0.8172

Table no 3: Fine-Tuning Phase – Accuracy and loss of training, validation and testing and its execution time per epoch

Model And Specifications			**Fine-Tuning Phase**						
Models	**Total Layers**	**Fine-Tuned Layers**	**Training Accuracy %**	**Validation Accuracy%**	**Training Loss**	**Validation Loss**	**Test Accuracy %**	**Test Loss**	**Training Time (secs)**
VGG19	22	5	98.74	96.90	0.034	0.0734	97.83	0.0603	1911.05
DenseNet121	427	86	96.08	97.52	0.1032	0.0826	96.90	0.0743	811.58
InceptionV3	311	63	97.61	97.21	0.0726	0.0838	96.59	0.0893	1213.02
MobileNetV3Large	187	38	82.67	65.63	0.5041	0.8376	79.26	0.6487	594.88

Table no 4: Transformer (ViT) Model Specifications and Training Time

Model	Total Parameters (M)	Transformer Blocks	Trainable Blocks (FT)	Frozen Blocks (FT)	Total Training Time (sec)
ViT (Base)	0.40	12	3	9	1262.53

Table no 5: Performance evaluation of the Vision Transformer (ViT) during Transfer learning and Fine-tuning.

Stage	Training Accuracy (%)	Validation Accuracy (%)	Train Loss	Validation Loss	Test Accuracy (%)	Test Loss	Time (secs)
Stage-1 (Transfer Learning)	49.87	68.11	0.8994	0.8924	–	–	868.80
Stage-2 (Fine-Tuning)	48.34	72.76	0.8965	0.8901	71.83	0.8896	393.73

4. Experiments and Results

Various experiments were conducted to train and evaluate the proposed deep learning framework for potato leaf disease classification. The objective of these experiments was to analyze the effectiveness of convolutional neural network (CNN) architectures and a Vision Transformer (ViT) model in identifying different classes of potato leaf diseases. The experiments were performed using the PlantVillage dataset, which contains labelled images of healthy and diseased plant leaves. Multiple pre-trained CNN models and a transformer-based architecture were trained using a transfer learning strategy followed by fine-tuning to improve classification performance.

The experimental pipeline involved dataset preparation, data augmentation, transfer learning, fine-tuning, and final testing. Performance was evaluated using accuracy and loss metrics across training, validation, and test sets. The experiments demonstrated that deep transfer learning models were effective for plant disease recognition, with improved accuracy after fine-tuning.

4.1 Experimental Setup

The experiments were carried out using an openly accessible dataset from PlantVillage. The dataset consists of 2,152 potato leaf images belonging to three categories: Early Blight, Late Blight, and Healthy samples. The images were organized into class-wise folders and then divided into training, validation, and test sets using a randomized splitting strategy to ensure a balanced class distribution across all subsets. The dataset was split in the ratio of 70% for training, 15% for validation, and 15% for testing. The training set contained the majority of the images and was used for model learning, while the validation set was used to monitor performance and tune hyperparameters during training. The test set was kept completely unseen during training and was used only for final performance evaluation. This randomized splitting approach ensured that the classes were proportionally distributed across all subsets, thereby improving the generalization ability of the models and preventing bias toward any particular class.

4.1.1 Hardware and Software Environment

All experiments were conducted on the Kaggle cloud computing platform using a single NVIDIA Tesla P100 GPU. The GPU acceleration enabled efficient training of deep neural network and transformer models. The implementation environment consisted of:

- Programming Language: Python

- Deep Learning Framework: TensorFlow with Keras API
- Transformer Library: Hugging Face Transformers (for ViT)
- Data Processing Libraries: NumPy, Pandas
- Visualization Libraries: Matplotlib, Seaborn

The experiments were executed in a controlled environment with fixed random seeds for Python, NumPy, and TensorFlow to ensure reproducibility of results.

4.1.2 Hyperparameters

The selection of hyperparameters was carefully designed to balance computational efficiency and classification performance within the Kaggle environment. All models were trained using a batch size of 32, which was chosen based on GPU memory limitations and to ensure stable gradient updates. Input images were resized to 224 × 224 × 3, ensuring compatibility with pre-trained ImageNet architectures such as VGG19, DenseNet121, InceptionV3, MobileNetV3Large, and Vision Transformer (ViT).For optimization, the Adam optimizer was used across all models. The initial learning rate was set to 1e-4 during Stage-1 (transfer learning). During fine-tuning (Stage-2), the learning rate was reduced to 1e-5 to allow gradual weight updates in the unfrozen layers and prevent catastrophic forgetting. The loss function used was categorical cross-entropy, appropriate for multi-class classification. Accuracy was used as the primary evaluation metric. Early stopping, model checkpointing, and learning rate scheduling were employed to improve convergence and prevent overfitting.

4.2 Training Process

The training process was divided into two stages:

Stage-1: Transfer Learning

In this stage, the convolutional or transformer backbone of each pre-trained model was frozen, and only the newly added fully connected classification head was trained. This allowed the models to leverage ImageNet-learned features while adapting to the PlantVillage dataset. Models evaluated in this stage include:VGG19, DenseNet121, InceptionV3, MobileNetV3Large, Vision Transformer (ViT). For CNN-based architectures, all convolutional layers were frozen during Stage-1. For ViT, all 12 transformer blocks were frozen initially.

Stage-2: Fine-Tuning

In Stage-2, the top 20% of layers (or transformer blocks in ViT) were unfrozen. This allowed high-level features to adapt specifically to potato leaf disease patterns. Fine-tuning significantly improved performance for several models as shown in table 6:

Table no 6: Transfer Vs Fine-tuned accuracy comparison for all models

Model	Transfer Accuracy	Fine-Tuned Accuracy	Gain
VGG19	0.9164	0.9690	+0.0526
ViT	0.6811	0.7276	+0.0464
InceptionV3	0.9474	0.9721	+0.0248
DenseNet121	0.9659	0.9752	+0.0093
MobileNetV3Large	0.6904	0.6563	-0.0341

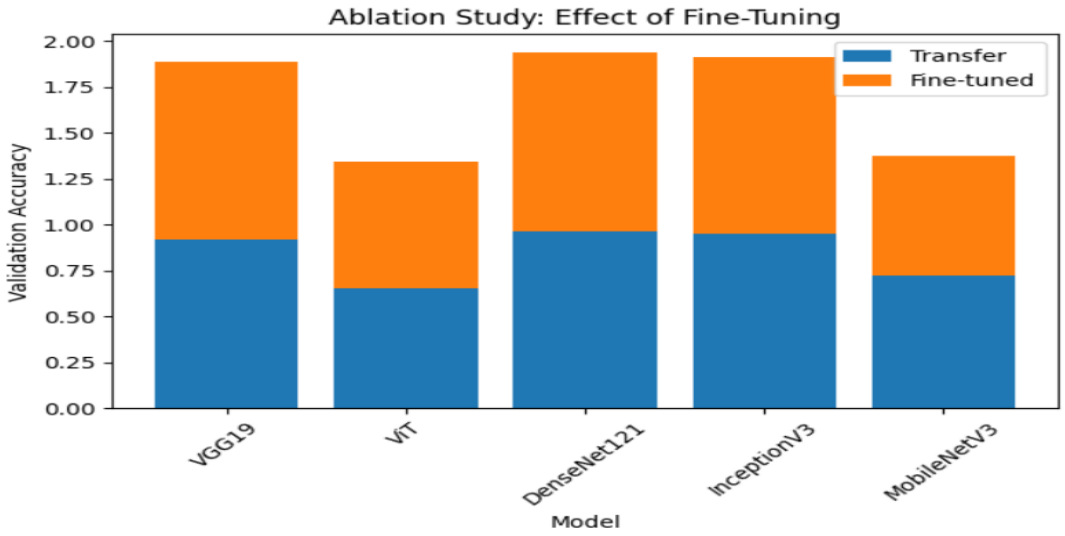

Figure 5: Impact of fine-tuning on validation accuracy across different models.

4.2.1 Batch Size and Epochs

All models were trained for a maximum of 100 epochs. However, training was controlled using:

- EarlyStopping (patience = 5)
- ReduceLROnPlateau (patience = 3)

This prevented unnecessary computation and reduced overfitting.

The Custom CNN achieved its best validation accuracy (0.9969) at epoch 83, demonstrating strong learning capability.

4.2.2 Learning Rate Scheduling

Learning rate scheduling was implemented using ReduceLROnPlateau:

- Initial learning rate: 1e-4
- Reduction factor: 0.2
- Patience: 3 epochs
- Minimum learning rate: 1e-6

This strategy allowed the model to make large updates initially and progressively refine weights during later epochs. The scheduling mechanism contributed to improved convergence stability, particularly for VGG19 and InceptionV3.

4.2.3 Accuracy on Test Set

The performance of all models on the unseen test dataset is summarized in Table 7. Among the evaluated architectures, VGG19 achieved the highest test accuracy (0.9783), followed by DenseNet121 (0.9690) and InceptionV3 (0.9659). The Custom CNN demonstrated competitive performance with an accuracy of 0.9257. In contrast, MobileNetV3Large and Vision Transformer (ViT) showed comparatively lower performance under the given dataset size.

Table no 7: Performance comparison of all models using accuracy, precision, recall, and F1-score.

Model	Accuracy	Precision	Recall	F1-Score
VGG19	0.9783	0.9419	0.9721	0.9559
DenseNet121	0.9690	0.9522	0.9409	0.9463
InceptionV3	0.9659	0.9289	0.9387	0.9337
Custom CNN	0.9257	0.9348	0.8485	0.8806
ViT	0.7276	0.4987	0.5244	0.5034
MobileNetV3Large	0.7926	0.5424	0.5688	0.5472

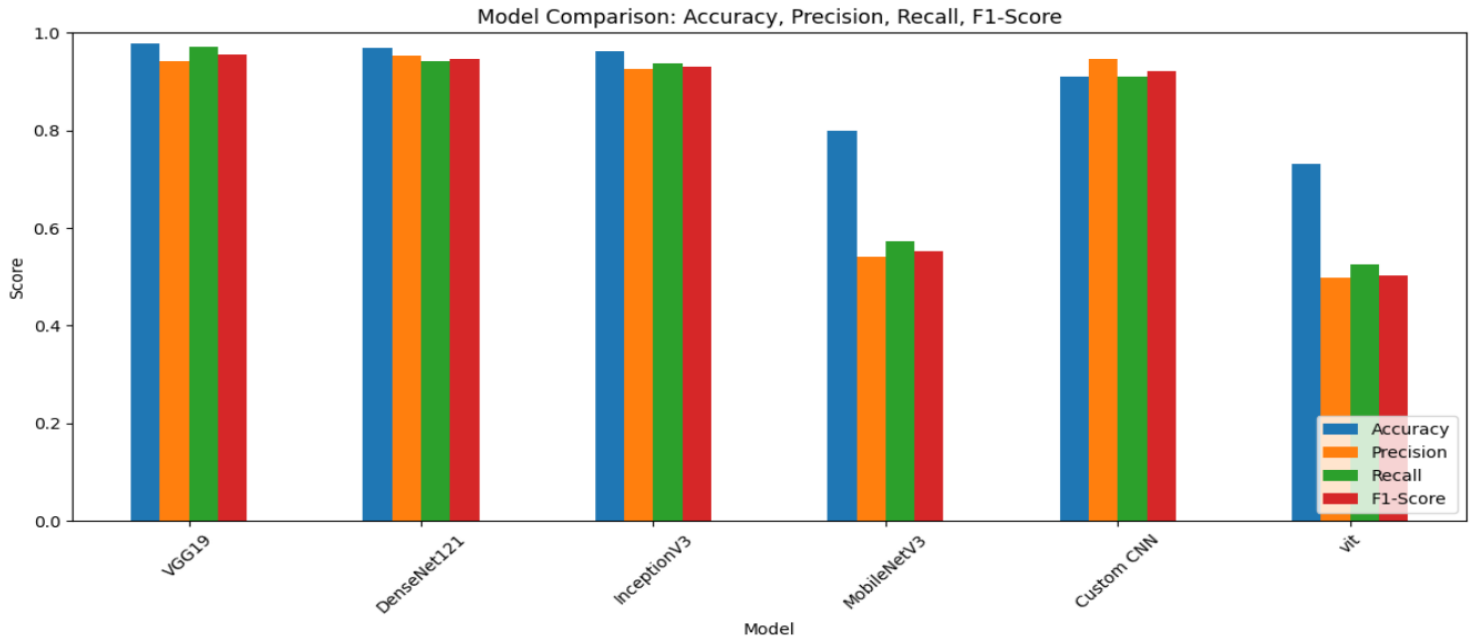

Figure 6: Bar Chart of Performance comparison of all models using Accuracy, Precision, Recall, and F1-score.

Although the Custom CNN achieved the highest validation accuracy (0.9969), VGG19 demonstrated superior overall generalization performance with balanced precision and recall.

4.4 Training and Validation Curve

The training and validation curves provide important insights into the learning behavior, convergence speed, and generalization capability of each model. In this study, training and validation accuracy and loss were monitored across epochs for all deep learning architectures, including the Custom CNN, VGG19, DenseNet121, InceptionV3, MobileNetV3Large, and Vision Transformer (ViT). The Custom CNN demonstrated a steady and consistent learning pattern throughout the training process. Both the training and validation accuracy curves increased gradually with each epoch, indicating stable learning and effective feature extraction from the dataset. The validation accuracy peaked at 0.9969 at epoch 83, showing that the model required more epochs to reach optimal performance. The training and validation loss curves decreased smoothly without significant fluctuations, which suggests that the model was not overfitting and maintained good generalization across the dataset. The VGG19 model exhibited very fast convergence compared to other architectures. Its validation accuracy reached 0.9814 by epoch 9, after which the performance stabilized. The training and validation curves for VGG19 were closely aligned, indicating minimal overfitting and strong generalization. The early convergence of VGG19 suggests that the pre-trained weights from ImageNet provided highly transferable features for the target dataset. DenseNet121 also showed strong and stable performance, reaching a validation accuracy of 0.9752 at epoch 6. The training and validation curves were smooth and closely matched, which reflects efficient feature reuse and strong gradient flow due to its dense connectivity structure. This architecture demonstrated fast convergence with high accuracy, making it one of the most effective pre-trained models in this study. InceptionV3 achieved a validation accuracy of 0.9721 at epoch 19. The training curves indicated gradual learning, and the validation accuracy improved steadily until convergence. The gap between training and validation accuracy remained small, suggesting that the model was able to generalize well without severe overfitting. MobileNetV3Large showed moderate performance, reaching a validation accuracy of 0.7492 at epoch 10. The training and validation curves exhibited fluctuations, indicating instability during training. This may be due to the lightweight architecture, which has fewer parameters and may not capture complex disease patterns as effectively as deeper networks. The Vision Transformer (ViT) achieved a validation accuracy of 0.7276 at epoch 17. The training curves indicated slower convergence compared to convolutional models. The gap between training and validation accuracy suggested limited generalization. This behavior is expected because transformer-based models typically require much larger datasets to perform optimally. Overall, the training and validation curves indicate that deeper convolutional architectures such as VGG19, DenseNet121, and InceptionV3 provided stable convergence and strong generalization. The Custom CNN

achieved the highest validation accuracy but required more epochs to reach optimal performance. Lightweight and transformer-based models demonstrated lower accuracy and less stable training behavior.

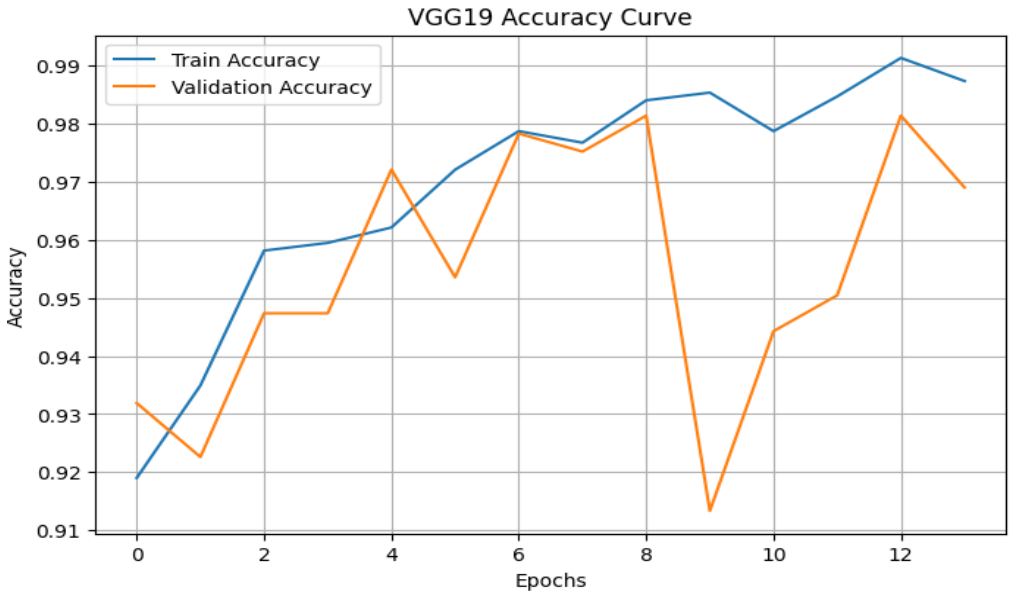

Figure 7: Training and Validation Accuracy over Epochs for VGG19 Model

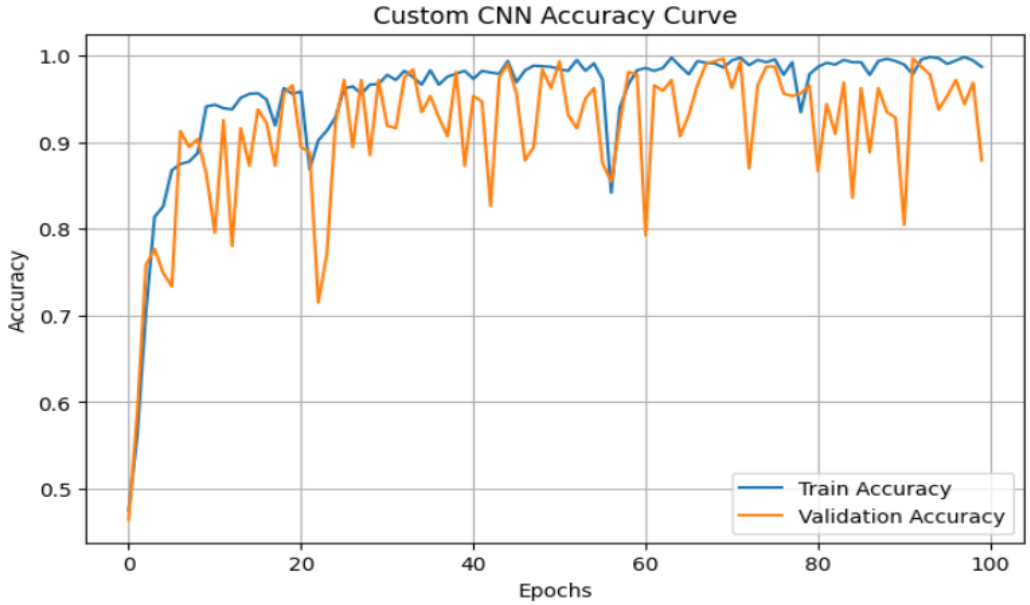

Figure 8: Training and Validation Accuracy over Epochs for Custom CNN Model

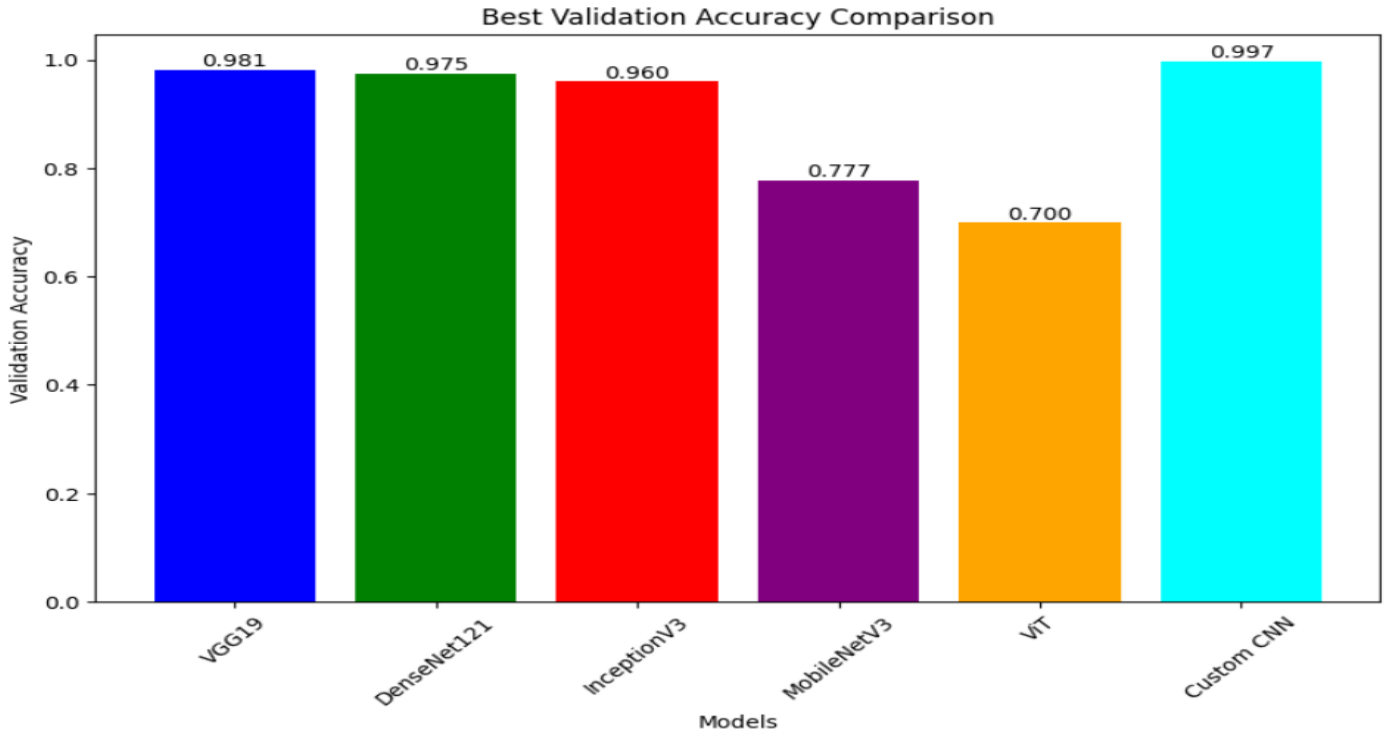

Figure 9: Bar Chart for Best Validation Accuracy for all models

5. DISCUSSION

The experimental results demonstrate significant performance variations across different deep learning architectures for potato leaf disease classification. The models were evaluated using accuracy, precision, recall, and F1-score to provide a comprehensive comparison of their classification capabilities. Among all the models, VGG19 achieved the highest overall performance, with an accuracy of 0.9783, precision of 0.9419, recall of 0.9721, and F1-score of 0.9559. These results indicate strong classification capability and balanced performance across all evaluation metrics. The success of VGG19 can be attributed to its deep convolutional architecture and effective transfer learning from pre-trained ImageNet weights, which provided robust feature representations. DenseNet121 and InceptionV3 also demonstrated strong performance. DenseNet121 achieved an accuracy of 0.9690 with an F1-score of 0.9463, while InceptionV3 achieved an accuracy of 0.9659 with an F1-score of 0.9337. The dense connectivity of DenseNet121 allowed efficient feature reuse and improved gradient flow, contributing to its high accuracy. InceptionV3 benefited from multi-scale feature extraction, enabling it to capture both fine and coarse disease patterns. The Custom CNN achieved an accuracy of 0.9257 and an F1-score of 0.8806. Although it achieved the highest validation accuracy during training, its overall test performance was slightly lower than the pre-trained models. This difference suggests that transfer learning provided stronger generalization compared to a network trained from scratch. However, the Custom CNN still demonstrated competitive performance and required no pre-trained weights, which may be beneficial in resource-constrained or domain-specific applications. MobileNetV3Large and Vision Transformer (ViT) showed moderate to low performance. MobileNetV3Large achieved an accuracy of 0.7926 with an F1-score of 0.5472, indicating limited capability in capturing complex disease patterns. The Vision Transformer (ViT) achieved an accuracy of 0.7276, along with relatively lower precision, recall, and F1-score compared to deep convolutional architectures, suggesting that transformer-based models require larger datasets for optimal performance. The lower performance of ViT is likely due to the relatively small dataset, as transformer-based models typically require large-scale training data. In summary, the results indicate that transfer learning with deep convolutional networks provides the most reliable performance for potato leaf disease classification. VGG19 emerged as the best-performing model, followed closely by DenseNet121 and InceptionV3. The Custom CNN also achieved high accuracy but required more training epochs. Lightweight and transformer-based architectures showed lower performance, highlighting the importance of model selection based on dataset size and complexity.

Table no 8: Classification Report for VGG19

	Precision	Recall	F1-Score	Support
Potato___Early_blight	0.99	0.99	0.99	150
Potato___Late_blight	0.97	0.99	0.98	150
Potato___healthy	0.95	0.91	0.93	23
Accuracy			0.98	323
Macro average	0.97	0.96	0.97	323
Weighted Average	0.98	0.98	0.98	323

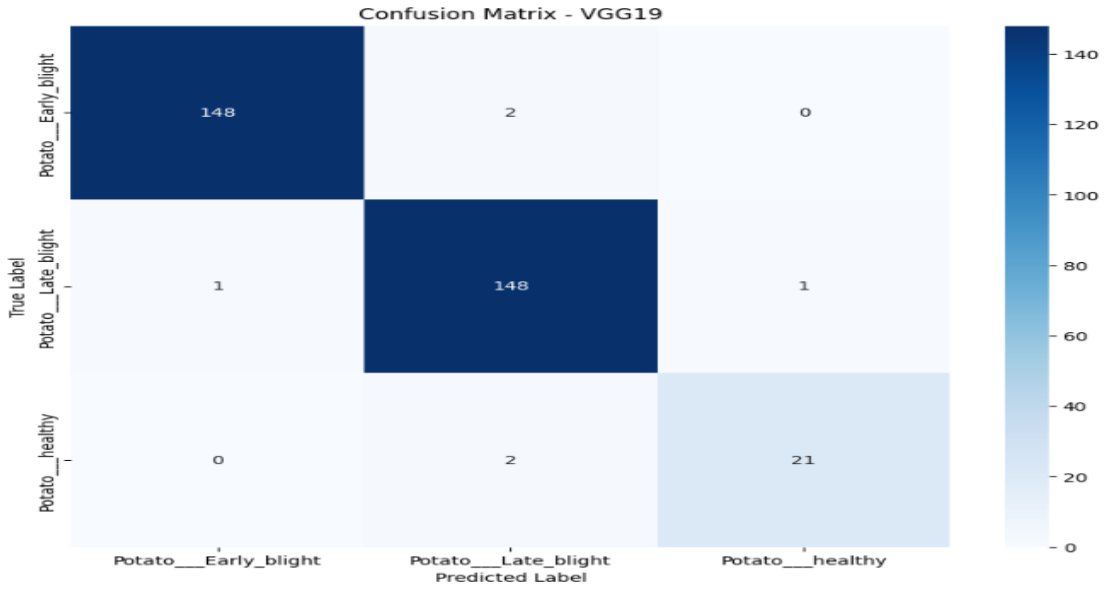

Figure 10: Confusion Matrix for VGG19

Table no 9: Classification Report for Custom CNN

	Precision	Recall	F1-Score	Support
Potato___Early_blight	0.88	1.00	0.94	150
Potato___Late_blight	1.00	0.89	0.94	150
Potato___healthy	0.95	0.83	0.88	23
Accuracy			0.93	323
Macro average	0.94	0.90	0.92	323
Weighted Average	0.94	0.93	0.93	323

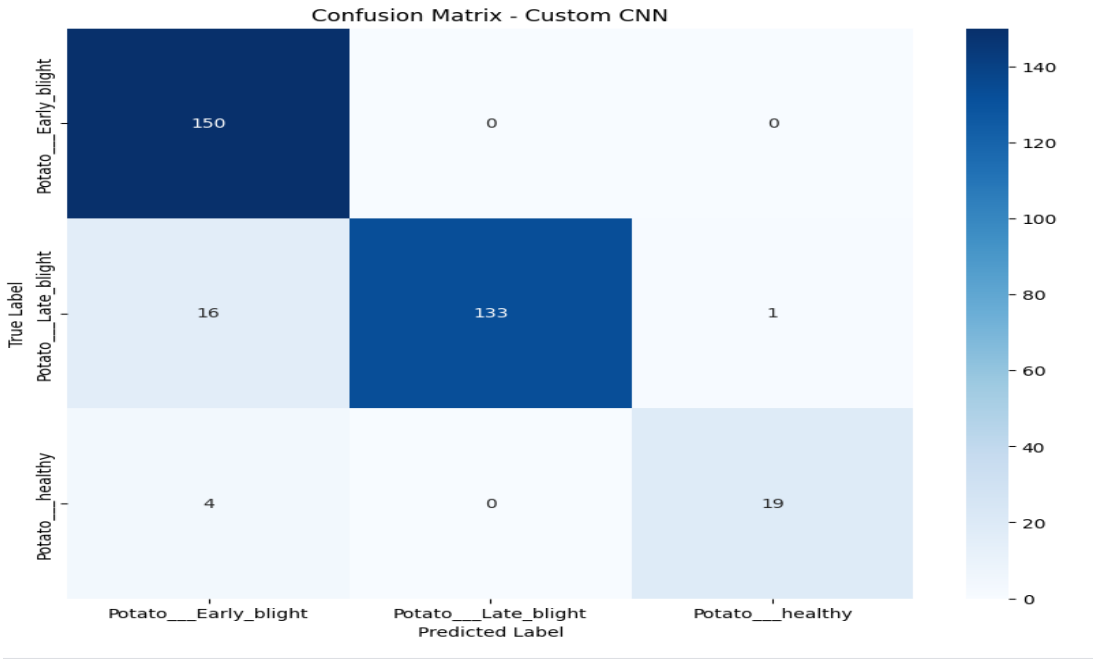

Figure 11: Confusion Matrix for Custom CNN

6. CONCLUSIONS

In this study, transfer learning and fine-tuning of state-of-the-art deep learning architectures were performed for image-based potato leaf disease classification. The models evaluated include VGG19, InceptionV3, DenseNet121, MobileNetV3Large, Vision Transformer (ViT), and a Custom CNN trained from scratch. A two-stage training strategy was adopted, where initially the backbone networks were frozen to leverage pre-trained ImageNet features, followed by fine-tuning of the upper layers using a reduced learning rate to enhance domain-specific feature adaptation. From the experimental analysis, deep convolutional architectures demonstrated superior learning capability compared to lightweight and transformer-based models under the given dataset size. Among all the evaluated models, VGG19 achieved the highest overall test accuracy of 97.83%, along with strong precision, recall, and F1-score values, indicating robust generalization and balanced classification performance across Early Blight, Late Blight, and Healthy classes. DenseNet121 and InceptionV3 also showed competitive performance, with stable convergence behavior and high classification accuracy. The Custom CNN achieved the highest validation accuracy during training; however, its test performance was slightly lower than the best transfer learning models, suggesting mild overfitting despite regularization strategies. The results further indicate that transfer learning significantly improves convergence speed and generalization compared to training from scratch. Fine-tuning of higher layers contributed to noticeable performance gains in several architectures, particularly VGG19 and InceptionV3, by allowing adaptation of high-level semantic features to disease-specific patterns. The Vision Transformer exhibited moderate performance, likely due to the relatively limited dataset size.

Overall, VGG19 proved to be the most reliable and effective architecture for multi-class potato leaf disease classification in this study. Although the achieved performance is highly satisfactory, future research can focus on reducing computational time, exploring larger and more diverse datasets, integrating advanced data augmentation strategies, and investigating hybrid CNN–Transformer architectures to further enhance robustness and real-world applicability in precision agriculture systems.

REFERENCES

1. D. Suo-meng and Z. Shao-qun, “Potato late blight caused by Phytophthora infestans: From molecular interactions to integrated management strategies,” Journal of Integrative Agriculture, vol. 21, no. 12, pp. 3456-3466, December 2022.
2. E. C. Too, L. Yujian, S. Njuki and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease,” Computers and Electronics in Agriculture, vol. 161, pp. 272-279, 2019.

3. S. Sauda, L. Kaur and M. Lal, "Comparison of Pre-trained Deep Model Using Tomato Leaf Disease Classification System," in Machine Intelligence and Smart Systems, 2021, pp. 553-566.
4. U. Y. Tambe, A. Shobanadevi, A. Shanthin and H.-. C. Hsu, "Potato Leaf Disease Classification using Deep Learning: A Convolutional Neural Network Approach," November 2023. [Online]. Available: https://arxiv.org/abs/2311.02338.
5. A. SOHEL, M. S. SHAKIL, S. M. T. SIDDIQUEE, A. A. MAROUF, J. G. ROKNE and R. ALHAJJ, "Enhanced Potato Pest Identification: A Deep Learning Approach for Identifying Potato Pests," IEEE Access, vol. 12, pp. 172149-172161, November 2024.
6. N. .H. Shabrina, S. Indarti, R. Maharani, D. A. Kristiyanti, I. N. Prastomo and T. A. M, "A Novel Dataset of Potato Leaf Disease in Uncontrolled Environment," Data in Brief, vol. 52, December 2023.
7. P. Mhala, A. Bilandani and S. Sharma, "Enhancing crop productivity with fined-tuned deep convolution neural network for Potato leaf disease detection," Expert Systems with Applications, vol. 267, 2025.
8. S. Chowdhury and D. K. Das, "Harnessing the Potato leaf disease detection process through proposed Conv2D and resnet50 deep learning models," Procedia Computer Science, pp. 539-547, 2025.
9. H. Gandhi, M. Lal and K. P. S. Attwal, "Advancements In Deep Learning Techniques For Image-Based Detection Of Diseases In Leaves Of Potato: A Review," International Organization of Scientific Research, vol. 28, no. 1, pp. 62-71, January 2026.
10. T. A. Dame, G. B. Adera and D. W. Girmaw, "Deep learning-based potato leaf disease classification and severity assessment," Discover Applied Sciences, vol. 7, June 2025.
11. A. T. Mulugeta, W. Jifara, E. Bogale, T. Desiyo and A. Mokonnen, "Early Detection and Classification of Potato Leaf Disease Using Convolutional Neural Networks," Applied Computational Intelligence and Soft Computing, vol. 2025, no. 1, November 2025.
12. G. Sangar and V. Rajasekar, "Potato Leaf Disease Classification using Pre Trained Deep Learning Techniques - A Comparative Analysis," International Conference for Emerging Technology (INCET), pp. 1-6, 2024.
13. A. Sharma, V. Singh, C. Sharma, G. Ansari, S. Kumar and K. Joshi, "Potato Leaf Disease Detection and Classification Using Deep Learning Technique," International Conference on Technological Advancements in Computational Sciences, pp. 2021-2030, January 2025.
14. S. M. Alhammad, D. S. Khafaga, W. M. El-hady, F. M. Samy and K. M. Hosny, "Deep learning and explainable AI for classification of potato leaf diseases.," Frontiers in Artificial Intelligence, vol. 7, February 2025.
15. E. Jain and A. Rathour, "Deep Learning-Based Detection of Potato Leaf Diseases Using VGG16: Early Blight, Late Blight, and Healthy Classification," International Conference on Computing Technologies, pp. 1-5, August 2025.