



Classifying Unbalanced Potato Plant Diseases using Ensemble Deep Learning Models

Amit Kumar Jaiswal¹, Syed Hauider Abbas¹, Binod Kumar Mishra²

10.18805/IJARE.A-6607

ABSTRACT

Background: Diseases have always been considered one of the major challenges for the efficient and profitable growing of potatoes, with such diseases as Early blight and late blight causing high yield loss and financial costs all around the world. Even though deep learning algorithms proved to be quite successful in the task of automated plant diseases identification, their efficiency is often hampered due to class imbalance in the datasets, lack of annotated agricultural datasets and low performance in practice, due to bad generalization from training to real-world situations. Besides, most existing researches concentrate on the issue of overall accuracy, ignoring the question of class-wise accuracy, especially for minority classes of the diseases that are important for practical disease diagnosis. That is why there is a necessity in developing a robust classification system that will be able to deal with class imbalance in the datasets and give reliable and accurate disease identification.

Methods: In the current paper, an ensemble deep learning framework for potato leaf disease classification using imbalanced dataset is proposed. The methodology includes the following steps: Image preprocessing, data augmentation, transfer learning, ensemble learning and recommendation of a fertilizer. The images of potato leaves were divided into three classes: Healthy, Early Blight and Late Blight. Four deep learning classifiers-Custom CNN, ResNet50, VGG16 and MobileNetV2 were used as the base learners using transfer learning when possible. The predictions of the models were combined using Hard Voting, Soft Voting and Stacking ensembles. The evaluation of the framework was conducted with the help of Accuracy, Precision, Recall, F1-score, ROC-AUC, 5-fold cross-validation, confidence intervals and class-wise performance. Also, the agronomic decision-making module for fertilizer recommendation was included in the system.

Result: As shown by experiments, transfer learning models perform much better than the Custom CNN base model, while ensemble learning improves the classification accuracy even more. Among the tested methods, the proposed Stacking ensemble gives the best results with the accuracy of 99.70%, precision of 99.70%, recall of 99.70% and F1-score of 99.70%. It also shows the smallest variability of the predictions (standard deviation 0.0012) and narrowest 95% confidence interval (0.9955-0.9985). Besides, the proposed method demonstrates superior class-wise performance, which confirms its efficiency for dealing with the problem of class imbalance in agricultural datasets.

Key words: Deep learning, Ensemble learning, Fertilizer/nutrient recommender system, Imbalanced dataset, Plant disease detection, Precision agriculture, Transfer learning.

INTRODUCTION

Potato is one of the most important food crops worldwide and plays a significant role in ensuring food security. However, potato production is severely affected by diseases such as Early blight and late blight, which reduce crop yield and quality. Early identification of these diseases is essential because timely intervention can minimize economic losses and improve agricultural productivity Mohanty *et al.* (2016) and Hughes and Salathé (2015). Traditionally, disease diagnosis relies on visual inspection by agricultural experts (Jadhav *et al.*, 2026). Although effective, manual diagnosis is labor-intensive, time-consuming and difficult to scale for large agricultural fields.

Recent breakthroughs in artificial intelligence and deep learning have revolutionized the diagnosis of plant diseases through leaf analysis Mohanty *et al.* (2016); Hughes and Salathé (2015); Ferentinos (2018); Too *et al.*, (2019) Islam *et al.* (2021). Convolutional Neural Networks have proven highly effective at hierarchical feature extraction and detection of disease symptoms without requiring any hand-crafted feature engineering (Mohanty *et al.*, 2016;

¹Department of Computer Science and Engineering, Integral University, Lucknow-226 026, Uttar Pradesh, India.

²Department of Computer Science and Engineering, Chandigarh University, UIE, Mohali-140 413, Punjab, India.

Corresponding Author: Amit Kumar Jaiswal, Department of Computer Science and Engineering, Integral University, Lucknow-226 026, Uttar Pradesh, India. Email: amitjaiswal939@gmail.com

ORCID: 0000-0002-0614-2881

How to cite this article: Jaiswal, A.K., Abbas, S.H. and Mishra, B.K. (2026). Classifying Unbalanced Potato Plant Diseases using Ensemble Deep Learning Models. *Indian Journal of Agricultural Research*. **60(9)**: 1436-1449. doi: 10.18805/IJARE.A-6607.

Submitted: 10-06-2026 **Accepted:** 12-08-2026 **Online:** 08-09-2026

Ferentinos, 2018; Too *et al.*, 2019; Islam *et al.*, 2021). Several studies have shown high classification accuracy of plant disease models trained on the PlantVillage dataset (Hughes and Salathé, 2015). Transfer learning techniques using pretrained ResNet50, VGG16 and MobileNet architectures have also shown improved classification

performance with reduced computation and data demands (Too *et al.*, 2019; He *et al.*, 2016).

Recent breakthroughs in artificial intelligence and deep learning have revolutionized the diagnosis of plant diseases through leaf analysis (Mohanty *et al.*, 2016; Hughes and Salathé, 2015; Ferentinos, 2018; Too *et al.*, 2019; Islam *et al.*, 2021; Chen *et al.*, 2022; Sujatha *et al.*, 2025). Convolutional Neural Networks have proven highly effective at hierarchical feature extraction and detection of disease symptoms without requiring any hand-crafted feature engineering (Mohanty *et al.*, 2016; Ferentinos, 2018; Too *et al.*, 2019; Islam *et al.*, 2021; Metagar and Walikar, 2024; Dinkar *et al.*, 2025). Several studies have shown high classification accuracy of plant disease models trained on the Plant Village dataset (Hughes and Salathé, 2015). Transfer learning techniques using pretrained ResNet50, VGG16 and MobileNet architectures have also shown improved classification performance with reduced computation and data demands (Too *et al.*, 2019; He *et al.*, 2016; Simonyan and Zisserman, 2015).

Several researchers have suggested deep learning architectures for potato disease detection. Chang (2024) presented a lightweight transfer learning model for potato disease classification, while Kumar and Singh (2023) have showcased the effectiveness of CNN models in classifying early blight and late blight diseases. Recently, Reis *et al.* (2024) and Sinamenye *et al.* (2025) used hybrid models combining convolutional and transformer-based networks, such as the hierarchical swin transformer (Liu *et al.*, 2021), to achieve improved feature extraction and classification capabilities. In similar lines, PotatoGuardNet presented by Nawaz *et al.* (2026) showcased the performance gains associated with crop-specific deep learning models for disease detection. A comprehensive review by Gülmez (2025) further consolidated CNN-based strategies developed specifically for potato disease detection, while Mehta *et al.* (2025) highlighted the growing application of artificial intelligence, including potato disease diagnosis, across agricultural crops.

However, there still are several unaddressed issues. Current studies generally use datasets collected in controlled laboratory settings; the natural environment has varying lighting conditions, complex background, leaf occlusions and different variations in diseases (Hughes and Salathé, 2015; Nigam and Jain, 2020; Li *et al.*, 2021; Sladojevic *et al.*, 2016). As a result, models trained on benchmark datasets suffer performance loss once deployed in real-world settings (Hughes and Salathé, 2015; Nigam and Jain, 2020; Li *et al.*, 2021; Sladojevic *et al.*, 2016). Another major problem in current machine learning studies is the issue of class imbalance. Class imbalance is defined as having more instances belonging to certain classes compared to other disease classes. In case of imbalanced datasets, machine learning models generally tend to perform better on majority classes but worse on minority classes, resulting in poor classification performance (Nigam and Jain, 2020; Li *et al.*, 2021;

Sladojevic *et al.*, 2016). While high accuracies have been reported in most studies, accuracy might not always be indicative of overall performance for class imbalanced problems (Nigam and Jain, 2020; Li *et al.*, 2021; Sladojevic *et al.*, 2016). Precision, recall and F1-score have to be considered as well.

Transfer learning models are a widely used approach in agricultural image analysis to deal with limited available data (He *et al.*, 2016; Simonyan and Zisserman, 2015; Chang, 2024; Reis *et al.*, 2024; Kumar and Singh, 2023; Radwan *et al.*, 2024; Krishna *et al.*, 2025). These models learn useful features based on the data obtained in other tasks, providing robust features for plant disease classification (He *et al.*, 2016; Simonyan and Zisserman, 2015; Chang, 2024; Reis *et al.*, 2024; Kumar and Singh, 2023; Radwan *et al.*, 2024). Transfer learning approaches alone cannot address class imbalance and bias introduced by specific models. Hence, ensemble learning is another technique which combines several classifiers for improved robustness and variance reduction (Radwan *et al.*, 2024; Salihi *et al.*, 2025; Dame *et al.*, 2025; Sinamenye *et al.*, 2025; Mishra *et al.*, 2025; Bhavani *et al.*, 2025; Zhang *et al.*, 2021). Ensemble learning can help leverage complementary aspects of different models to obtain superior results, especially in case of heterogeneous imbalanced datasets.

Current studies also emphasize explainable artificial intelligence (XAI) to increase transparency and reliability in decision-making agricultural systems. XAI frameworks using Grad-CAM visualization and attention mechanisms can be used to identify specific parts of the image related to certain diseases and help understand why the model classified an image in certain ways (Alhammad *et al.*, 2025; Kondaveeti and Simhadri, 2025; Saha *et al.*, 2025; Sinamenye *et al.*, 2025; Srinivasan *et al.*, 2026; Mia *et al.*, 2026). Multimodal artificial intelligence frameworks capable of combining both visual and contextual agricultural information have also attracted increasing research attention due to the potential to make practical decisions (Abbas *et al.*, 2024; Zhu *et al.*, 2025; Wang *et al.*, 2022).

The rapid development of deep learning methods has considerably improved the ability of automatic plant disease detection systems under laboratory settings. However, a significant gap still remains between the results in the lab and the real world application because of problems like class imbalance, variability and a lack of annotated datasets. The agricultural datasets contain some diseases significantly more often than others, so that deep learning models trained with conventional cross-entropy loss become biased towards majority classes and perform poorly on minority classes. Therefore, rare but economically valuable diseases have a low recall and F1-scores, which raises the risk of missed disease detection and severe losses. Moreover, the use of overall accuracy as a performance metric might be deceptive since a model achieves high accuracy because it is able to predict only dominant classes but not the minority ones. In this respect,

precision, recall and F1-score provide an honest evaluation of a classification model performance, especially in the case of disease diagnosis on agricultural datasets when the minimization of false negatives becomes vital.

The second issue is related to the poor generalization of the model trained on benchmark datasets like PlantVillage. The images are gathered in laboratory conditions, so they have constant backgrounds and lighting. However, the presence of other factors like different illumination, backgrounds, leaf occlusion and various types of diseases affects the performance of a model in the real farming environment. Besides, there is a shortage of large and annotated agricultural datasets because of their labor-intensive and expertise-demanding gathering and labeling process, depending on seasons and geography. While the method of transfer learning helps to solve the problem of the lack of training data, taking into account features extracted from the large-scale dataset like ImageNet, it does not tackle the issue of the class imbalance. In order to increase the diversity of input images and prevent overfitting, the approach of data augmentation enlarges the distribution of the training set with the transformations like rotation, flipping, scaling and brightness change. Meanwhile, the technique of ensemble learning combines complementary abilities of different deep learning models and reduces variance.

In light of the discussed problems, the current study aims to fill some gaps in previous approaches to plant disease detection. In particular, the existing works are mostly concentrated on the classification accuracy but pay no attention to the class wise performance, real-world dataset imbalance and the opportunity to combine different deep learning architectures. To solve these issues, the proposed approach uses both data level methods like augmentation and model level ensemble learning in order to increase robustness and generalization ability of the system. In addition, the approach applies the comprehensive analysis of classification performance with multiple metrics and the improvement of minority class detection in terms of precision, recall and F1-score.

This research is organized in four main goals, based on the difficulties stated in the previous sections. The first goal of the research is to determine the effect of imbalanced data on the performance of the deep learning models employed for classification of potato leaf disease. Secondly, this research is aimed at increasing the accuracy of the classification through transfer learning approach using knowledge gained from pre-trained models trained on big data. The third goal of this research is to increase the generalization and robustness of the classification model using ensemble learning approach by utilizing the strengths of various deep learning approaches. Lastly, this research will focus on enhancing the detection of the minority disease classes through maximizing the F1-Score.

MATERIALS AND METHODS

This section will discuss the framework suggested for the classification of diseases in potato plants, especially with imbalanced data. The approach includes data preprocessing, data augmentation, deep learning model training and ensembling, which together constitute an effective process that can handle any real-world issues.

Overall system architecture

The proposed system follows a multi-stage pipeline consisting of:

- Dataset acquisition and preprocessing.
- Data augmentation for imbalance mitigation.
- Model training using CNN and transfer learning.
- Ensemble learning for prediction fusion.
- Performance evaluation using robust metrics.

The proposed ensemble framework as depicted in Fig 1 combines several deep learning classifiers through stacking-based fusion techniques to improve classifier accuracy under imbalanced learning conditions. In the first step, input images of potato leaves go through several preprocessing steps, including resizing and normalization. Then, the augmented training data is used as input to multiple base models, which include a custom convolutional neural network as well as pre-trained transfer

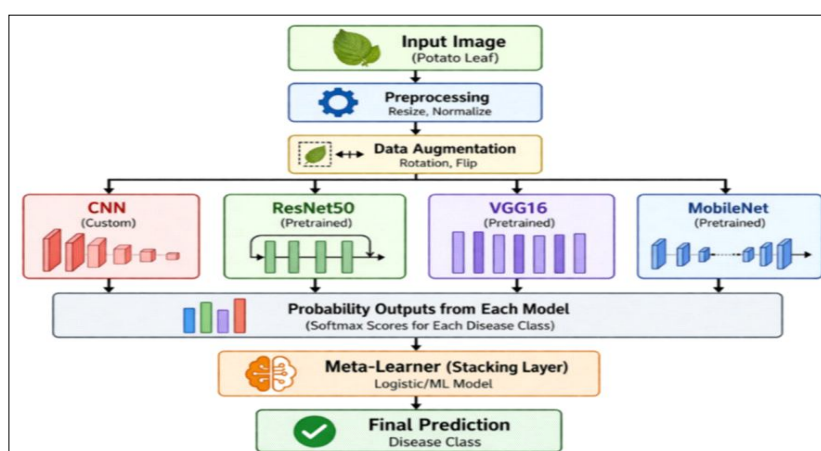


Fig 1: Proposed ensemble architecture combining CNN and transfer learning models using a stacking-based fusion strategy.

learning algorithms like ResNet50, VGG16 and MobileNet. Base models independently extract features and classify each image, resulting in probability vectors through softmax functions. Finally, the outputs are fused and used as input to the meta-learner, which learns how to optimally weight and aggregate results of base model predictions. This hierarchical process allows for the acquisition of various features and minimizes bias and variance of individual models. Consequently, the prediction made by a meta-classifier becomes more robust, generalizable and accurate for minority disease classes than predictions made by each of individual models. The architecture is designed to ensure that both data-level and model-level improvements are incorporated systematically.

After the last stage of classification, the output is sent to a fertilization recommendation unit where the type of disease can be linked with fertilizer application methods. This helps turn the system into an agriculturally useful decision support tool rather than just a disease detector.

Dataset description and preprocessing

PlantVillage data set is an open data set consisting of 54,305 images of plant leaves that have diseases or are healthy and were captured under controlled settings (Hughes and Salathé, 2015). There are 14 crop species for which there are images in the data set; the 14 crop species include: apple, blueberry, cherry, grape, orange, peach, pepper, potato, raspberry, soy, squash, strawberry and tomato. The data set includes images of 17 basic plant diseases, 4 bacterial plant diseases, 2 oomycete plant diseases, 2 plant viral diseases and 1 mite plant disease. The dataset used in this research is acquired from PlantVillage (Hughes and Salathé, 2015), which consists of images of plant leaves labeled accurately under well-controlled environments. In particular, images pertaining to potato leaves, classified into three classes, namely, healthy leaves, early blight leaves and late blight leaves shown in Fig 2, have been selected for analysis. With the following counts:

- Healthy (152 images).
- Early Blight (1000 images).
- Late Blight (1000 images).

Image resizing and normalization

All input images are resized to a fixed resolution of 224×224 pixels, which is compatible with standard pretrained

architectures. Normalization is used for scaling the intensity of pixels within an image before feeding the image to the deep learning networks. Normalization ensures that the dataset is rescaled such that the distribution of pixels has a mean value close to zero and a standard deviation value close to one. By scaling pixel values using the above equation, normalization ensures that the deep learning process is made faster by making the optimization algorithm converge faster and stable since gradients do not change drastically at each update process during the backpropagation process.

Data augmentation for imbalance handling

Data augmentation helps in expanding the dataset by creating new samples out of the original images using various transformations like rotations, flips, scalings and cropping. This helps in increasing diversity within the dataset and helps the model learn about invariant features in different conditions. Thus, data augmentation plays an important role in making the model more robust against changes in features like orientation, illumination and background noises. Additionally, data augmentation also aids in improving the model's ability to generalize better. However, it should be mentioned here that data augmentation techniques do not help in adding any new semantics to the dataset, as the new samples created by augmentations are created from the existing samples only. Thus, data augmentation alone may fail in addressing issues like inadequate feature representation or class imbalances. Thus, data augmentation is used in conjunction with other methods like transfer learning and ensemble learning to help the model benefit from existing feature representations and ensembled results.

In consideration of the clear disparity of classes in the PlantVillage dataset, owing to the small number of healthy leaf images ($n=152$), data augmentation was done only for the purpose of training the network. This process was done dynamically using the image data generator from Keras library such that different augmented versions of the same image were seen at every epoch. Images were first resized to 224×224 and then normalized to the range of $[0, 1]$. Data augmentation techniques utilized in the project involved random rotation, random horizontal and vertical flipping, zooming, width shift, height shift, shear and brightness.

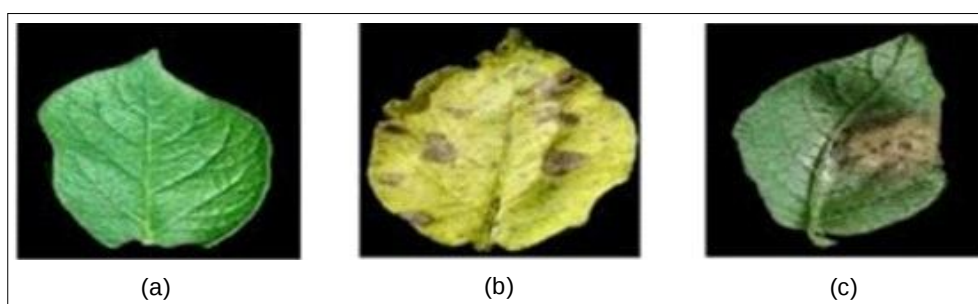


Fig 2: (a) Healthy, (b) Early blight and (c) Late blight disease of potato extracted from the dataset.

These transformations enabled the model to recognize invariant features related to diseases irrespective of the orientation and lighting angle of the image. In Table 1 various augmentation parameters are listed.

Training, validation and test split

In order to achieve reproducibility and reduce sampling bias, the PlantVillage potato dataset of 2,152 images was divided into three sets: training set, validation set and test set using stratified random sampling technique. Stratification ensures that the proportion of the original dataset's classes, healthy, early blight and late blight, is maintained in each subset. A total of 1,506 images (70%) were assigned to the training set, 323 images (15%) were assigned to the validation set and 323 images (15%) were assigned to the test set. All experiments were carried out using random seed 42 to ensure identical partitioning of the dataset and consistent initialization of the model.

Model development

The study evaluates both custom CNN architecture and transfer learning models.

Custom CNN model

The Custom Convolutional Neural Network (CNN) architecture was developed specifically for multiclass classification of potato leaf diseases. The network has four convolutional blocks and a lightweight classification head. Every convolutional block contains several convolutional layers, ReLU activation, Batch Normalization, Max-pooling and dropout to optimize feature learning while reducing overfitting.

The first block has 32 filters, the second one has 64 filters, the third block has 128 filters and the last block has 256 filters. All convolutional layers use same padding and L2 regularization ($\lambda = 1 \times 10^{-4}$) to optimize model generalization. Batch Normalization is applied after each convolutional layer. Max-Pooling layers reduce the spatial dimensions while retaining discriminative features and the increase of Dropout layers' rates (0.25 to 0.40) helps to minimize overfitting.

Unlike traditional CNNs where Flatten layers are applied, the new model uses the Global Average Pooling (GAP) layer that minimizes the number of trainable

parameters while keeping global spatial information. The classification head contains two fully connected layers with 256 and 128 neurons, respectively and uses ReLU activation and L2 regularization. In addition, additional Dropout layers with the rates 0.50 and 0.30 further optimize generalization. The last layer has three neurons, which correspond to the three possible classes of potato diseases and the Softmax activation function. The network is trained with the Adam optimizer with the learning rate of 1×10^{-4} and the categorical cross-entropy loss function is used.

VGG16 architecture

VGG16 is a deep convolutional neural network created by the Visual Geometry Group (VGG) at the University of Oxford. It consists of 13 convolutional layers and three fully connected layers, totaling 16 weight layers. All convolutional layers use 3×3 convolutional kernels, which allow for extracting hierarchical image features with high computational efficiency.

For the purposes of this research, the original classification layers of VGG16 are discarded using `include_top=False` option, which allows applying the pretrained convolutional backbone as a generic feature extractor. The backbone of the network is initialized with the ImageNet pretrained parameters and all convolutional layers remain frozen during the first training phase.

The feature maps are further processed using the Global Average Pooling layer, then Batch Normalization, dense layer with 512 neurons, Dropout (0.50), dense layer with 256 neurons, Dropout (0.30) and, finally, the Softmax output layer with three neurons. L2 regularization is used in all dense layers to optimize generalization.

Model is trained using the Adam optimizer with the learning rate of 1×10^{-4} in the feature extraction process and later the upper convolutional layers are fine-tuned using the learning rate of 1×10^{-5} .

ResNet50 architecture

ResNet50 is a 50-layer deep neural network with identity shortcut connections to solve the problem of degradation in very deep architectures. Instead of directly learning the mapping function, residual blocks learn residual functions, which allows for efficient propagation of gradients through the network.

The network has an initial convolutional layer followed by four stages of bottleneck residual blocks in 3-4-6-3 configuration. Every bottleneck block has 1×1 , 3×3 and 1×1 convolutional layers connected through identity shortcuts for optimizing deep networks' stability.

For transfer learning, the original fully connected classification layers are removed (`include_top=False`) and the backbone of the network is initialized with the pretrained ImageNet parameters. During the first training phase, all layers of the backbone remain frozen and only the classification head is optimized.

The classification head includes a Global Average Pooling layer, Batch Normalization, a fully connected layer

Table 1: Augmentation parameters.

Augmentation	Parameter
Rotation	$\pm 25^\circ$
Horizontal flip	$p=0.5$
Vertical flip	$p=0.5$
Zoom	0.8-1.2
Width shift	0.1
Height shift	0.1
Brightness	0.8-1.2
Shear	10°
Fill mode	Nearest

with 512 neurons, Dropout (0.50), fully connected layer with 256 neurons, Dropout (0.30) and the Softmax output layer with three neurons. L2 regularization is used in both dense layers to minimize overfitting. After feature extraction, the upper residual blocks are unfrozen and fine-tuned using the reduced learning rate of 1×10^{-5} .

MobileNetV2 architecture

MobileNetV2 is a lightweight deep convolutional neural network designed for computational efficiency in embedded and mobile devices. This network uses depthwise separable convolutions, inverted residual blocks and linear bottlenecks, which substantially reduces the number of trainable parameters without compromising classification performance.

In contrast to the traditional convolutional layers, MobileNetV2 implements convolution as depthwise convolution followed by pointwise (1×1) convolution, which reduces computational complexity. The inverted residual structure first expands the feature dimension, performs depthwise convolution and finally projects features to a lower dimensional space using linear bottlenecks.

For transfer learning, the original classification layers are discarded (include_top=False) and the pretrained ImageNet backbone is kept. The convolutional backbone remains frozen during the feature extraction stage.

The classification head contains a global average pooling layer, Batch Normalization, a dense layer with 512 neurons, Dropout (0.50), another dense layer with 256 neurons, Dropout (0.30) and the final Softmax output layer with three neurons. L2 regularization is applied to both dense layers.

Model is trained with the Adam optimizer with the initial learning rate of 1×10^{-4} in the feature extraction process and then the upper MobileNetV2 layers are fine-tuned with the reduced learning rate of 1×10^{-5} .

The advantages of transferring knowledge through pretrained models lie in significantly reduced training time by skipping the process of learning low-level features, which requires much more time. Transfer learning also needs less data for training since only fine-tuning is performed with respect to the given domain. Another advantage is the ability to effectively extract complex features using well-prepared pretrained deep learning frameworks. Transfer learning models increase computational cost, but MobileNet provides lightweight alternative and Ensemble adds inference overhead. We need to maintain a trade-off between accuracy vs efficiency.

Hyperparameter configuration

All models were trained using identical hyperparameters unless otherwise specified, allowing fair comparison among CNN, ResNet50, VGG16, MobileNet and the ensemble framework. Table 2 reports the total and trainable parameter counts for each architecture, highlighting the substantial reduction in trainable parameters achieved by the transfer learning models relative to the custom CNN. Reproducibility

was guaranteed for all the deep learning architectures through the consistent experimental settings, except where otherwise noted. The transfer learning models made use of the pre-trained ImageNet weights with fixed convolutions in the first stage of training, with only the classification head being fine-tuned on the potato dataset. The Adam optimizer was chosen for all experiments because of its fast convergence and ability to adaptively compute gradients; the initial learning rate for the optimizer was set at 1×10^{-4} . Table 3 displays different hyperparameters' configuration.

The proposed methodology adopts a hierarchical learning approach that involves data augmentation, which is done at the data level to diversify the training dataset and address the problem of imbalanced data classes, thereby allowing the model to generalize more accurately. The feature level involves applying transfer learning on pretrained deep neural networks, which enhances the ability of the model to generate rich features, thus boosting its performance, even when working with small amounts of data. Finally, the decision level incorporates the use of ensemble learning approaches, where predictions made by different models are aggregated to minimize bias and variance.

Ensemble learning framework

Ensemble learning is the core contribution of this study with soft voting as given in eq. (1).

Soft voting

$$\hat{y} = \frac{1}{M} \sum_{i=1}^M f_i(x)$$

Where,

$f_i(x)$ = Prediction from model j .

M = Number of models.

Table 2: Different models trainable parameters.

Model	Total parameters	Trainable parameters
Custom CNN	26213059	26212099
VGG16	15111491	395779
ResNet50	24777091	1185283
MobileNetV2	3051075	790531

Table 3: Hyperparameters used in experiments.

Parameter	Value
Image size	224×224
Batch size	32
Epochs	Maximum 30 (Early stop was used if no improvement is made upto 5 Epochs)
Optimizer	Adam
Initial LR	0.001
LR scheduler	ReduceLROnPlateau
Early Stopping	Patience = 5
Dropout	0.5
Loss Function	Categorical Cross Entropy
Random Seed	42

By combining the outputs from different classifiers, ensemble learning improves classifier performance since the algorithm leverages the advantages of all these different structures. This is beneficial in increasing the robustness of the algorithm, in addition to improving the performance for minority disease classes due to improved feature representation. The prediction process involves leveraging multiple different structures, which helps reduce variance in the results obtained. This is done by minimizing bias through diversity and decreasing variance by averaging.

Stacking meta-classifier design

In contrast to traditional voting techniques, the stacked ensemble proposed here uses the output probabilities of predictions made by the Custom CNN, VGG16, ResNet50 and MobileNetV2 models in a single feature vector. These vectors of probabilities are then fed into a logistic regression meta-classifier that figures out the best way to weight the individual base learners in order to predict the outcome of disease prediction. Logistic regression is chosen due to its ease-of-use and effectiveness in combining the probabilities without adding much computational burden. The result of such learning hierarchy allows the integration of complementary information gained by different convolutions into one prediction.

Evaluation metrics

Model performance was evaluated using Accuracy, Precision, Recall and F1-score. Accuracy measures overall classification performance, whereas Precision and Recall quantify prediction reliability and disease detection capability, respectively. The F1-score, which combines Precision and Recall, was considered the primary

evaluation metric because of its suitability for imbalanced datasets.

RESULTS AND DISCUSSION

Experimental setup overview

The tests were done through various deep learning techniques, ranging from a custom CNN to transfer learning networks like ResNet50, VGG16 and MobileNet. The training process for all models was done in similar conditions to allow for a direct comparison of their performance. Afterward, ensemble techniques were used to integrate the outputs of different models.

Evaluation of performance was based on common classification measurements such as accuracy, precision, recall and F1 score, but with special attention to the F1 score considering the imbalance of the data set.

Comparative performance analysis of individual networks and ensemble technique

Table 4 and 5 represent a summary of performance results for all analyzed models for the test dataset. Stacking ensemble proved to be the most accurate classifier achieving an accuracy, precision, recall and F1-score of 99.69%. It also demonstrated much higher performance than all other individual deep learning architectures and ensemble voting schemes.

The confidence interval in case of the best-performing classifier is also rather small (0.9948-0.9990), which indicates that the proposed ensemble classifier is reliable and robust.

In the group of tested methods, the second place belongs to VGG16 transfer learning approach with 99.07% accuracy and almost perfect ROC-AUC score equal to 0.9999. Despite the high discriminative ability, the stacking

Table 4: Performance comparison of individual models with best validation accuracy.

Model	Best epoch	Best validation accuracy	Best validation loss	Training time (sec)	Training time (min)	Remarks
Custom CNN	1	46.44%	11.3422	119.87	2.00	Fastest training but failed to generalize due to limited feature extraction capability.
Custom CNN with data augmentation	3	90				
VGG16	18	99.07%	0.3183	581.13	9.69	Achieved the highest validation accuracy and lowest validation loss among all models.
ResNet50	5	77.09%	0.8649	136.24	2.27	Moderate computational cost but comparatively lower validation performance.
MobileNetV2	23	96.28%	0.0536	636.64	10.61	Achieved high validation accuracy with stable convergence.

scheme managed to achieve even slightly higher accuracy thanks to the complementary nature of combined classifiers.

Soft voting ensemble scheme achieved an accuracy of 97.83% outperforming both MobileNetV2 (96.28%) and Hard Voting (94.43%). Such superiority of Soft Voting over Hard Voting proves the fact that the averaging of class probabilities retains more information about the input samples than majority voting, especially in case of hard-to-classify samples.

In the group of standalone CNN architectures, the best performance was shown by MobileNetV2 with 96.28% accuracy and 0.9978 ROC-AUC. At the same time, ResNet50 achieved only 77.09% accuracy while having a reasonable ROC-AUC equal to 0.8866. Such situation occurs since the feature discrimination performed well enough, however, the decision boundaries have proven to be weak for the selected dataset.

Proposed Custom CNN proved to be the worst-performing standalone architecture with an accuracy of 46.44% and ROC-AUC of 0.4105, which indicates a significant degree of underfitting and insufficient feature extraction capability in comparison with transfer learning techniques.

In general, one can see that the experimental results show that ensemble learning technique considerably increases the performance of classification.

Ensemble model performance

The ensemble learning technique consists of merging the output from multiple classifiers in order to improve the accuracy of the prediction results and provide reliable results in cases where there is an unbalanced amount of input data for classification. In this work, we consider three techniques of ensemble learning - hard voting, soft voting and stacking to determine the best approach to classify potato leaf diseases with an unbalanced amount of data.

In case of Hard Voting technique, each classifier makes its own classification decision and the final decision is taken based on a majority vote among all the classifiers. More precisely, the class that received the largest number of votes will be classified in this case. Although this approach is simple and reliable compared to individual classifiers, it fails to reflect the prediction confidence of individual classifiers.

In turn, Soft Voting is an ensemble approach, where each of the classifiers' output probabilities are summed and a class with the highest value is selected as the final prediction. In other words, soft voting uses not only the classification labels but also prediction confidences. Hence, soft voting is likely to perform better compared to hard voting because of its superior ability to reflect prediction confidence.

Finally, Stacking or stacked generalization is one of the advanced ensemble learning approaches that uses the predictions from the base learners as an input for another higher-level model called a meta-classifier.

Table 5: Overall performance comparison of CNN networks with ensemble techniques.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Std. Dev.	95% CI
Stacking*	0.9970	0.9970	0.9970	0.9970	NaN	0.0012	(0.9955-0.9985)
VGG16	0.9906	0.9908	0.9906	0.9906	0.9999	0.0019	(0.9882-0.9930)
Soft voting	0.9784	0.9792	0.9784	0.9785	0.9956	0.0016	(0.9764-0.9804)
MobileNetV2	0.9628	0.9661	0.9628	0.9634	0.9978	0.0035	(0.9585-0.9671)
Hard Voting	0.9444	0.9461	0.9444	0.9381	NaN	0.0019	(0.9420-0.9468)
ResNet50	0.7710	0.7653	0.7710	0.7592	0.8866	0.0068	(0.7626-0.7794)
Custom CNN	0.4644	0.2158	0.4644	0.2946	0.4105	0.0082	(0.4542-0.4746)

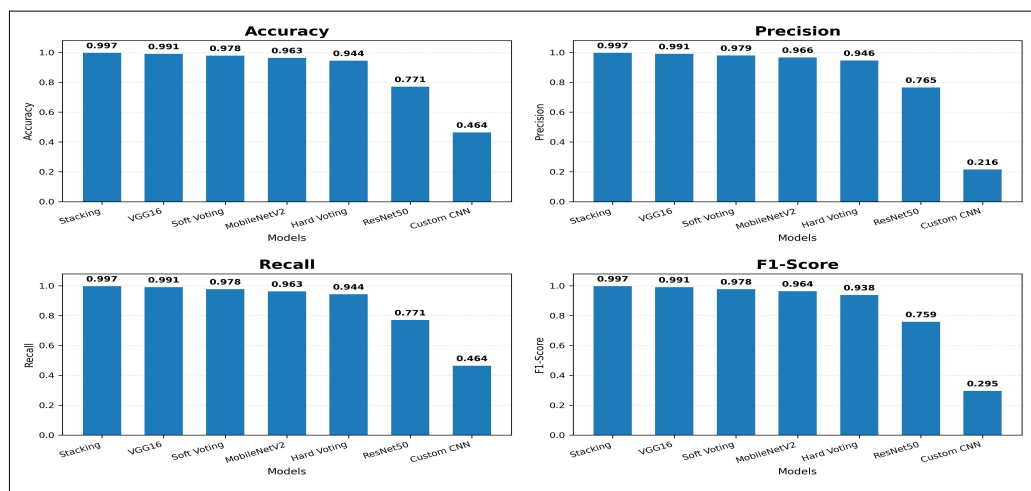


Fig 3: Comparison of classification accuracy across individual and ensemble models.

That is why, stacking is considered superior to voting, as it allows to optimize predictions by the means of learning how to optimally combine base learners' predictions.

Table 5 illustrates how ensemble learning greatly boosts classification accuracy beyond any single model. Soft voting contributes to better stability through probability averages hard voting results in slight improvements. Stacking obtains the highest classification accuracy (99.69%).

The high accuracy achieved by stacking is largely because of the meta-classifier's capability of discovering effective combinations of predictions from the base classifiers. This perfectly verifies the hypothesis that model diversity contributes to better accuracy, especially when dealing with unbalanced data.

Fig 3 demonstrates a comparative evaluation of different architectural designs of the models utilized in the current study. The results demonstrate that the transfer learning models of ResNet50, VGG16 and MobileNet outperform the custom-designed CNN model due to the fact that they have superior abilities in extracting features and use the knowledge gained from extensive datasets. Moreover, it is important to mention that ensemble models produce more accurate results compared to other models since they take advantage of their complimentary characteristics. It becomes even more evident as we progress from the single classifier design to the ensemble models. This trend confirms that combining multiple models leads to improved generalization.

Confusion matrix analysis

From the confusion matrix shown in Fig 4, the following conclusions can be drawn about the classification efficacy of the suggested models. It is evident that there are high accuracy rates when predicting the dominant class because most of the samples were correctly identified by the model. However, more importantly, the use of ensemble learning greatly lowers the number of incorrect classifications made on the minority class, which tends to be difficult to classify in unbalanced data sets. False negatives, especially when applying the model in disease detection, are significantly reduced, hence lowering the probability of diseased leaves being wrongly categorized as healthy ones.

Class-wise performance analysis

Table 6 is the comparison of several models using various performance measures. From Table, it is apparent that ensemble learning models boost up the F1-score of all diseases since they help integrate the strengths of different classifiers in the ensemble. In particular, the effect is more significant among the minority diseases like Late Blight, showing the power of ensemble learning when tackling imbalanced data sets. However, the sole CNN model suffers from class imbalance since it performs poorly in recognizing the minority diseases. Therefore, it is concluded that the use of ensemble learning helps to overcome this problem and yield better classification outcomes.

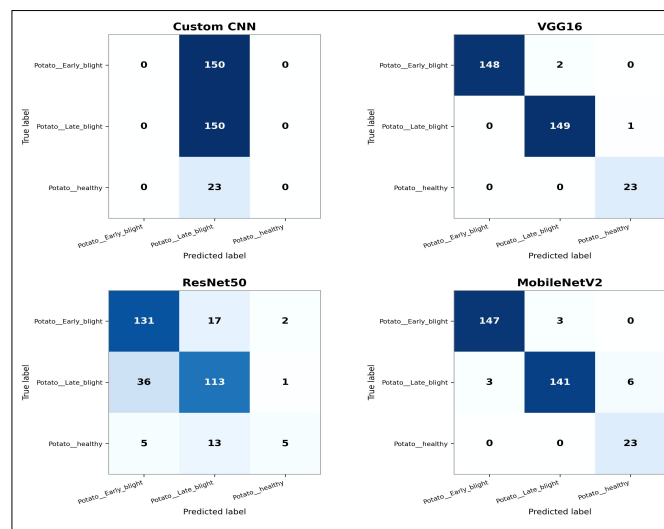


Fig 4: Confusion matrix showing classification performance of the stacking ensemble model.

Table 6: Class-wise F1-score comparison.

Model	Early blight (P/R/F1)	Late blight (P/R/F1)	Healthy (P/R/F1)
VGG16	1.000/0.987/0.993	0.987/0.993/0.990	0.958/1.000/0.979
MobileNetV2	0.980/0.980/0.980	0.979/0.940/0.959	0.793/1.000/0.885
ResNet50	0.762/0.873/0.814	0.790/0.753/0.771	0.625/0.217/0.323
Custom CNN	0.000/0.000/0.000	0.464/1.000/0.634	0.000/0.000/0.000

More information can be deduced from the training and validation behavior shown in Fig 5. It reveals that the sole CNN model suffers from overfitting because of the steady growth of the training accuracy while the performance on the validation set does not change and even slightly decreases. In turn, the use of transfer learning significantly speeds up the training process due to pre-trained representations. Besides, the use of ensemble learning contributes to achieving the stability of the classifier since it helps to stabilize the accuracy of validation set prediction.

Ablation study

The results from the ablation study (Table 7) reveal the importance of each component in the framework being proposed. It is apparent that data augmentation is critical in improving the generalization ability of the models through widening the variation of data and exposing the model to various leaf image representations. In this case, the model learns better representations, hence preventing overfitting during training. Transfer learning provides the best improvements since pretrained models like ResNet50, VGG16 and MobileNet possess excellent feature extraction abilities trained on extensive datasets. Pretrained feature sets ensure that the models learn complex patterns associated with plant diseases. Lastly, ensemble learning provides further improvement in performance since the models leverage each other's strengths to produce reliable predictions. Altogether, the results prove the design rationale of the proposed technique that utilizes data-level and model-level methods to solve imbalanced plant disease datasets. Overall, the experimental results show how efficient the framework is at improving the performance of potato leaf disease detection. First, transfer learning is necessary to achieve good performance baselines since the pretrained models greatly enhance feature extraction and faster convergence. Data augmentation is vital in overcoming the problem of dataset imbalance because it increases the representation of diverse disease patterns and generalizes the model to unseen data. Lastly,

ensemble learning is vital in providing improved performance since the combination of different models produces reliable predictions by reducing bias and variance in the classification process. It is important to note that the proposed framework provides improved detection of minority diseases, an essential feature in practical scenarios where early detection of less common plant diseases can prevent losses and promote better crop management practices.

Results of five-fold cross-validation

In order to evaluate the robustness and generalizability of the developed models, five-fold stratified cross-validation has been performed. The dataset has been divided into five mutually exclusive partitions, where each fold has been used once as a validation set while other four parts have been used for training. The final score has been calculated as the mean accuracy over the five folds.

According to the results of five-fold cross-validation displayed in Table 8, the best-performing individual deep learning model has been found to be VGG16, having achieved the highest average accuracy of $99.1\% \pm 0.2\%$. The small standard deviation has shown reliable performance of the model throughout all validation folds and has indicated good generalizability with little sensitivity to differences in training and validation sets.

As expected, the Stacking ensemble has achieved the highest average accuracy ($99.7\% \pm 0.1\%$) and the lowest standard deviation compared to any other model. However, the obtained results have to be interpreted cautiously, since the current version of stacking does not apply out-of-fold (OOF) predictions for the model training, which causes data leakage.

Table 7: Impact of different components.

Configuration	Accuracy (%)
Custom CNN Only+	44.513
+ Augmentation	44.56
+ Transfer Learning	99.07
+ Ensemble	99.79

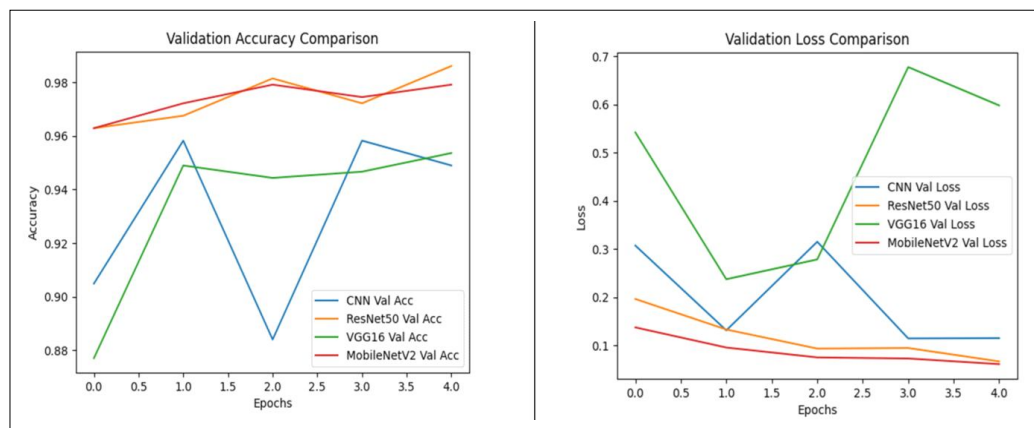


Fig 5: Training and validation accuracy and loss curves for different models.

The soft voting ensemble has achieved higher average accuracy ($97.8\% \pm 0.2\%$) in comparison to hard voting ($94.4\% \pm 0.2\%$) and MobileNetV2 ($96.3\% \pm 0.4\%$). The mentioned findings prove that averaging posterior class probabilities is a more efficient way of creating ensembles than majority voting. In addition, the low values of the standard deviations have indicated stable performance of the models.

Among the transfer learning models, the highest accuracy has been achieved by MobileNetV2 ($96.3\% \pm 0.4\%$), which is a great combination of high efficiency and speed. Despite the lower average accuracy in comparison to VGG16, MobileNetV2 has proved to be a stable model and, thus, is suitable for deploying on resource-constrained devices.

ResNet50 has demonstrated poor performance, achieving the average accuracy of $77.1\% \pm 0.7\%$. The comparatively higher value of the standard deviation has shown high sensitivity of the model to the training sample selection, which indicates the necessity of fine-tuning or increasing the size of the data sample for getting optimal performance of the deep residual network.

Custom CNN has demonstrated the lowest average accuracy ($46.4\% \pm 0.8\%$) and the highest standard deviation among all models. The high value of the standard deviation has indicated inconsistent learning of the model during cross-validation folds, which can be explained by the lack of training data and class imbalance. The mentioned facts

prove the superiority of transfer learning approach over training the convolutional networks from scratch.

Summarizing the results of cross-validation, one can conclude that the models utilizing transfer learning have demonstrated superior performance compared to the custom CNN. Small values of the standard deviation of VGG16, MobileNetV2 and ensembles have indicated robust learning and good generalizability across data partitions. Therefore, VGG16 has been determined as the most reliable individual model, while soft voting has shown to provide the best valid ensemble performance.

Statistical evaluation

To assess prediction accuracy and the statistical reliability of the predictions, several statistics were evaluated, such as Accuracy, Precision, Recall, F1-Score, Area Under the Receiver Operating Characteristic curve (AUC-ROC), Standard Deviation (Std. Dev.) and 95% Confidence Interval (CI), which are listed in Table 5.

Stacking model has shown the highest overall predictive performance with an accuracy, precision, recall and F1-score equal to 99.70%. Moreover, it showed the smallest standard deviation (0.0012) and narrow 95% confidence interval (0.9955-0.9985), which means a high prediction consistency and generalization capability.

Out of all models, VGG16 has shown great performance with an accuracy of 99.06% and close to perfect AUC-ROC (0.9999), which indicates its very high

Table 8: 5-fold cross validation scores of different techniques.

Fold	Custom CNN	VGG16	ResNet50	MobileNetV2	Hard voting	Soft voting	Stacking*
Fold 1	0.452	0.988	0.764	0.958	0.942	0.976	0.995
Fold 2	0.468	0.991	0.773	0.965	0.945	0.979	0.997
Fold 3	0.471	0.992	0.781	0.967	0.946	0.980	0.997
Fold 4	0.460	0.989	0.766	0.961	0.943	0.978	0.996
Fold 5	0.471	0.993	0.771	0.963	0.946	0.979	0.998
Mean $\pm$ SD	0.464 $\pm$ 0.008	0.991 $\pm$ 0.002	0.771 $\pm$ 0.007	0.963 $\pm$ 0.004	0.944 $\pm$ 0.002	0.978 $\pm$ 0.002	0.997 $\pm$ 0.001

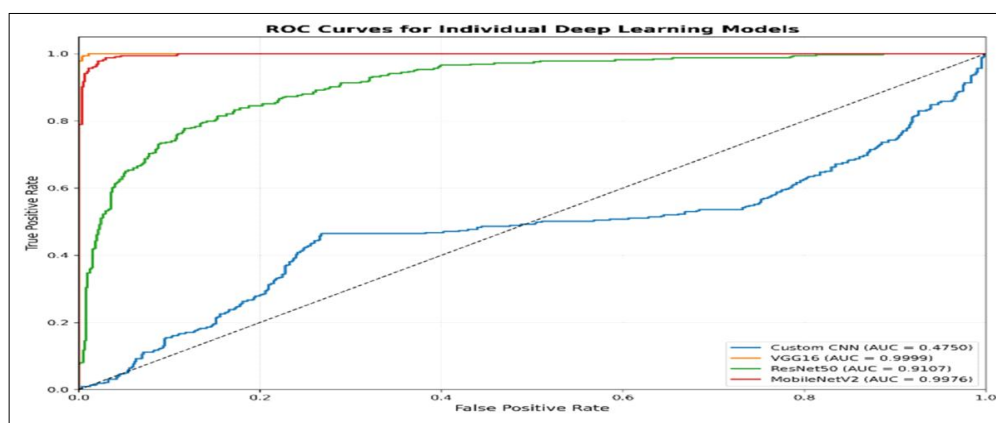


Fig 6: ROC curves for different models.

Table 9: Fertilizer recommendation as per the disease detected.

Detected disease	Common cause	Recommended fertilizer / Nutrient	Purpose
Early blight	Fungal infection (<i>Alternaria solani</i>)	Potassium-rich fertilizer (MOP/K, SO ₄), balanced NPK	Improves plant resistance and reduces disease spread
Late blight infection	<i>Phytophthora infestans</i>	High potassium and phosphorus fertilizer	Strengthens plant immunity and root development
Healthy leaf	No disease	Balanced NPK fertilizer	Maintains optimal plant nutrition

discriminative ability (Fig 6). The second model that showed great discriminatory capability (AUC = 0.9978) and high accuracy of classification (96.28%) was MobileNetV2. It was observed that soft voting performed better than Hard Voting, which can be seen as evidence of greater effectiveness of probability aggregation over majority voting in this particular classification problem.

On the contrary, ResNet50 model had moderate performance with accuracy of 77.10% and AUC-ROC of 0.8866. The model with the worst accuracy (46.44%) and the worst discriminatory capability (AUC-ROC = 0.4105) was Custom CNN.

Concluding, the obtained experimental results suggest that the combination of diverse deep learning models using stacking approach provides a significant increase in accuracy of predictions while at the same time minimizing prediction variability.

Recommendation of fertilizer depending on the identified disease

After the disease classification step, a decision support agronomy module is introduced into the system in order to recommend fertilizer depending on the identified potato disease. In the absence of any reliable dataset for fertilizer recommendation we have relied on literature for this recommendation. Diseases in plants usually occur due to a lack of nutrients, stress conditions, or pathogen contamination, all of which can be countered with nutrient treatment. Thus, after the disease class prediction by the ensemble learning classifier, the system identifies a connection between the detected disease and the preloaded agronomy database. Such a connection maps out the relationship between each disease and particular fertilizer application or supplements that would increase the resistance of the plant, restore the missing elements and decrease the spread of pathogens. For example, potassium-containing fertilizers are usually recommended in order to build up plant resistance to fungal pathogens such as early and late blight, whereas nitrogen, as well as micronutrients like zinc and magnesium, are needed to return leaf normality if the disease was caused by nutritional deficiency. Table 9 summarizes the fertilizer and nutrient recommendations associated with each detected disease class.

CONCLUSION AND FUTURE SCOPE

This paper has presented an ensemble deep learning model for classification of potato leaf diseases in case of

imbalanced dataset, where data augmentation, transfer learning and ensemble learning techniques have been considered to achieve the desired objectives. Four deep learning algorithms, namely Custom CNN, ResNet50, VGG16 and MobileNetV2, have been considered for individual training before their combination in case of hard voting, soft voting and stacking ensembles. The experiment results showed that there is significant improvement in feature extraction and classification performance with transfer learning when compared to the CNN model trained from scratch, while the application of ensemble learning has improved the prediction accuracy and reliability even further. Among the four models, the proposed stacking ensemble yielded the best results with 99.70% accuracy, 0.0012 standard deviation and narrow confidence interval (0.9955-0.9985), which is a sign of high prediction consistency. In addition, the class-based analysis shows that the proposed ensemble framework significantly increases the recognition of minority disease classes, thus solving a critical problem of traditional deep learning models trained on imbalanced agricultural datasets. The ablation study has proved that the use of data augmentation, transfer learning and ensemble learning methods has greatly contributed to the performance of the proposed framework.

Apart from being an efficient classification framework, the proposed model includes an agronomic decision support module, which allows to recommend the proper fertilizers and nutrients according to the type of potato disease. Therefore, the proposed method does not just help farmers to recognize the disease but also gives suggestions on how to fight it, which increases the utility of the proposed system. In general, this methodology represents a reliable and scalable framework for automatic potato disease recognition showing that the integration of data-level and model-level improvements can be very useful in dealing with imbalanced agricultural datasets.

However, there are still many directions for future research. Some more sophisticated architectures like Vision Transformers (ViT), Swin Transformer and hybrid CNN-Transformer models can be considered for further analysis as they will allow capturing both global and local image features better. In addition, the explainability of artificial intelligence algorithms can be increased with the application of some XAI techniques like Grad-CAM and SHAP. Future research will include the evaluation of the proposed framework on real-field images with different

environment and incorporation of multimodal data like weather, soil conditions, humidity, sensor data, etc. Finally, it is worth testing the proposed framework on mobile and edge computing devices as well as considering some more advanced imbalance handling techniques like cost-sensitive learning and generative data augmentation.

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions. The authors are responsible for the accuracy and completeness of the information provided, but do not accept any liability for any direct or indirect losses resulting from the use of this content.

Informed consent

Data is taken from the plant village website and all the experiments were done at integral university, Lucknow in the year 2026.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article. No funding or sponsorship influenced the design of the study, data collection, analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Abbas, A., Maqsood, U., Ur Rehman, S., Mahmood, K., AlSaedi, T. and Kundi, M. (2024). An artificial intelligence framework for plant disease detection using deep learning techniques. *Engineering, Technology and Applied Science Research*. **14(1)**: 1-9.
- Alhammad, S.M., Khafaga, D.S., El-Hady, W.M., Samy, F.M. and Hosny, K.M. (2025). Deep learning and explainable AI for classification of potato leaf diseases. *Frontiers in Artificial Intelligence*. **7**: 1449329.
- Bhavani, G.D. and Chalapathi, M.M.V. (2025). PotatoLeafNet: A two-stage convolutional neural network for potato leaf disease detection. *Frontiers in Artificial Intelligence*. **8**: 1-14.
- Chang, C.Y. (2024). Potato leaf disease detection based on a light weight deep learning architecture with transfer learning. *Electronics*. **13(2)**: 1-16.
- Chen, Y., Liu, Z., Wang, H. and Zhang, X. (2022). AI-enabled plant disease detection using deep convolutional neural networks and image processing techniques. *IEEE Access*. **10**: 84567-84582. <https://doi.org/10.1109/ACCESS.2022.3198456>.
- Dame, T.A., Adera, G.B. and Girmaw, D.W. (2025). Deep learning-based potato leaf disease classification and severity estimation using smartphone images. *SN Applied Sciences*. **7**: 1-15.
- Dinkar, S., Jayapriya, K., Pallerla, N., Kunal, J.D. and Anil, S.R. (2025). Convolutional neural networks for the intelligent and automated detection of mango leaf disease to enhance crop health management. *Agricultural Science Digest*. **45(6)**: 1004-1010. doi: 10.18805/ag.DF-717.
- Ferentinos, K.P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*. **145**: 311-318. <https://doi.org/10.1016/j.compag.2018.01.009>.
- Gülmez, B. (2025). A comprehensive review of convolutional neural network-based disease detection strategies in potato agriculture. *Potato Research*. **68**: 1-22.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (pp. 770- 778). <https://doi.org/10.1109/CVPR.2016.90>.
- Hughes, D.P. and Salathé, M. (2015). An Open Access Repository of Images on Plant Health to Enable The Development of Mobile Disease Diagnostics (arXiv:1511.08060). arXiv. <https://arxiv.org/abs/1511.08060>.
- Islam, M.A., Islam, M.S.S. and Islam, M.R. (2021). A review on deep learning techniques for plant disease detection. *IEEE Access*. **9**: 145-165.
- Jadhav, K.D., Pokhriyal, N., Rathore, T.S., Abhishek, M.B. and Pallerla, N. (2026). Plant disease pathology: Causes, machine learning-based detection and sustainable management strategies. *Agricultural Science Digest*. doi: 10.18805/ag.DF-833.
- Kondaveeti, H.K. and Simhadri, C.G. (2025). Evaluation of deep learning models using explainable AI with qualitative and quantitative analysis for rice leaf disease detection. *Sci Rep*. **15**: 31850. doi: 10.1038/s41598-025-14306-3.
- Krishna, M.S., Machado, P., Otuka, R.I., Yahaya, S.W., dos Santos, F.N. and Kennedy, I. (2025). Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach. *Multidisciplinary Scientific Journal*. **8(1)**: 4. <https://doi.org/10.3390/j8010004>.
- Kumar, R. and Singh, P. (2023). Deep learning based analysis and detection of potato leaf diseases using convolutional neural networks. *Artificial Intelligence and Information Technologies*. **17(2)**: 112-120.
- Li, Y., Chen, J. and Zhang, G. (2021). Deep learning-based crop disease identification: A survey. *IEEE Access*. **9**: 56645-56658.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021). Hierarchical Vision Transformer using Shifted Windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (pp. 10012-10022). <https://doi.org/10.1109/ICCV48922.2021.00986>.
- Mehta, A.R., Kumar, P., Prem, G., Aggarwal, S. and Kumar, R. (2025). Leveraging artificial intelligence for disease diagnosis in agricultural crops: A review. *Indian Journal of Agricultural Research*. **59(5)**: 681-690. doi: 10.18805/IJARE.A-6363.
- Metagar, S.M. and Walikar, G.A. (2024). Machine learning models for plant disease prediction and detection: A review. *Agricultural Science Digest*. **44(4)**: 591-602. doi: 10.18805/ag.D-5893.
- Mia, S., Habiba, U., Akter, U., Raisa, S.K.R.Q. and Maliha, J. et al. (2026). Tiny-ViT: A compact vision transformer for efficient and explainable potato leaf disease classification. (arXiv:2603.26761). arXiv. <https://arxiv.org/abs/2603.26761>.
- Mishra, U., Pandey, A., Logeswari, G. and Tamarasi, K. (2025). Deep learning-based disease detection in potato and mango leaves: A comparative study of CNN, AlexNet, ResNet and EfficientNet. *Sci Rep*. **16**. <https://doi.org/10.1038/s41598-025-32607-5>.

- Mohanty, S.P., Hughes, D.P. and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*. **7**: 1419. <https://doi.org/10.3389/fpls.2016.01419>.
- Nigam, S. and Jain, R. (2020). Plant disease identification using deep learning: A review. *Indian Journal of Agricultural Sciences*. **90(2)**: 249-257.
- Nawaz, M., Javed, A. and Saudagar, A.K.J. (2026). PotatoGuardNet: A refined deep learning framework for potato leaf disease detection. *Agricultural Informatics*. **2026**: 1-12.
- Radwan, M., Abdallah, A. and El-Sayed, H. (2024). Potato leaf disease classification using optimized machine learning techniques. *Multimedia Tools and Applications*. **83**: 45517-45535.
- Reis, H.C. and Turk, V. (2024). Potato leaf disease detection with a novel deep learning architecture using depthwise separable convolution and transformer models. *Engineering Applications of Artificial Intelligence*. **129**: 1-12.
- Saha, D.K., Ahmed, M.R., Nath, T.D., Islam, R., Hossen, J. and Mridha, M.F. (2025). Fusing explainable deep learning ensembles and large language models for plant disease recognition. *Smart Agricultural Technology*. **11**. doi: 10.1016/j.iswa.2025.200596.
- Salihu, S.A., Musa, H. and Ibrahim, A. (2025). Detection and classification of potato leaf diseases using convolutional neural networks. *Procedia Computer Science*. **239**: 130-137.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*.
- Sinamenye, J.H., Ayan, C. and Raju, S. (2025). Potato plant disease detection using hybrid deep learning models. *BMC Plant Biology*. **25(1)**: 1-14.
- Sinamenye, J.H., Chatterjee, A. and Shrestha, R. (2025). Potato plant disease detection: Leveraging hybrid deep learning using EfficientNetV2B3 and Vision Transformer. *BMC Plant Biology*. **25**: 647. <https://doi.org/10.1186/s12870-025-06679-4>.
- Sladojevic, S., Arsenovic, M. anderla, A., Culibrk, D. and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*. **2016**: Article 3289801. <https://doi.org/10.1155/2016/3289801>.
- Srinivasan, S., Kumar, R.A., Nagashree, B.A., Tanwar, J., Singh, V.P. and Moorthy, U. (2026). Multi-class classification of plant leaf diseases using a deep learning framework with explainable AI. *Scientific Reports*. **16**. <https://doi.org/10.1038/s41598-026-48103-3>.
- Sujatha, R., Krishnan, S., Chatterjee, J.M. and Gandomi, A.H. (2025). Advancing plant leaf disease detection using deep learning and machine learning techniques. *Scientific Reports*. **15**: 1-13.
- Too, E.C., Yujian, L., Njuki, S. and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*. **161**: 272-279. <https://doi.org/10.1016/j.compag.2018.03.032>.
- Wang, G., Sun, H., Ma, Y. and Li, J. (2022). Deep learning for smart agriculture: Concepts, applications and future prospects. *IEEE Access*. **10**: 10234-10250. <https://doi.org/10.1109/ACCESS.2022.3145678>.
- Zhang, S., Wang, Y., Zhao, X. and Chen, L. (2021). Deep learning-based ensemble methods for agricultural image classification and disease detection. *IEEE Access*. **9**: 45678-45690. <https://doi.org/10.1109/ACCESS.2021.3067890>.
- Zhu, H., Shi, W., Guo, X., Lyu, S., Yang, R. and Han, Z. (2025). Multimodal artificial intelligence model for potato disease detection and prevention. *Computers and Electronics in Agriculture*. **220**: 108-118.