Early Detection and Classification of Potato Leaf Disease Using an Efficient Deep Learning Model

Author(s): **Samiksha Madan More, R Dileep Kumar, Hari Mohan Singh**

Affiliation(s): Department of Computer Science and Information Technology,
Vaugh Institute of Agricultural Engineering And Technology,
Sam Higginbottom University of Agriculture, Technology And Sciences, Prayagraj,
U.P, INDIA 211007

Corresponding Author: sameekshamorey@gmail.com,

Abstract:

Plant diseases impact the availability and safety of plants for human and animal consumption, as well as the safety of food, limiting food availability and access and crop output and quality. Especially plants that contribute to the national economy as well as the food needs of its own people. Manual diagnosis of disease in early stages requires so much of expertise, technology assisted automatic disease classification approaches are becoming new norm now a days. In this study, an effort is made to detect potato leaf disease as India is second largest country producing potato. In this research a novel deep learning model for potato leaf disease classification has developed using CNN models and tested on PlantVillage dataset. The proposed model is trained to accurately classify potato leaf disease in 3 classes & the method achieved 99.61% accuracy. The outcome of the experiment demonstrates that, the proposed model outperforms pre-trained model i.e. VGG16 and InceptionV3. The proposed method also tested with respect to its consistency and reliability.

Keywords: potato leaf disease classification, deep learning, crop health, convolutional neural network, supervised learning

1. Introduction

Across the globe, potato is the fourth largest agriculture food crop after rice, wheat and maize. India produces 48.5 million tons of potatoes annually and rank 2nd largest country in the production of potato. According to the Agricultural and Processed Food Products Export Development Authority (APEDA), Uttar Pradesh is the nation's top producer of potatoes, contributing more than 30.33% of the nation's total output. Annual yield of crop heavily depends on various issues like climatic conditions, diseases, soil health etc. Among these, Plant diseases has adverse role in crop yield if not diagnosed and treated in time. It damage the leaves, stems, roots and fruits and have an impact on crop quality and quantity, which results in global food insecurity and famine [1]. Around 16% of worldwide crop yields are projected to be lost each year as a result of crop disease, which is the main factor behind food scarcity and rising food production costs [2].The Food and Agriculture Organization (FAO) predicts that by 2050, there will be about 9.1 billion people on the planet. Approximately 70% of food production increase is necessary for a consistent supply of food [3].

Diseases and disorders are the several types of causes that have an impact on plants and the things they produce. Diseases brought on by algae, fungi, or bacteria are considered biotic factors, whereas rainfall, moisture, temperature, and nutrient deprivation are considered biotic factors that cause diseases [4]. The appropriate diagnosis of Potato leaf diseases and timely remedial action can save the potato from wastage. Moreover it can ensure the good quality of the potato yield that eventually gives maximum profit to farmers.

There are multiple techniques to identify plant diseases but one of the straight forward approaches is a visual estimation. The traditional techniques for identifying plant diseases rely on farmers experience and expertise, which can be unpredictable and unstable. Researchers have developed a spectrometer to distinguish between healthy and diseased plant leaves in place of more traditional methods for identifying plant diseases [5].Such methods are expensive &time consuming as they demand a highly professional operation, experiment conditions, and extensive use of crop protection agents. The development of automated plant leaf disease detection methods is now made possible thanks to recent advancements in artificial intelligence (AI), machine learning (ML), Deep Learning (DL) and computer vision (CV) technology. These methods can quickly and precisely identify plant leaf diseases without the need for human intervention. It has been noted that agriculture has used DL the most frequently.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

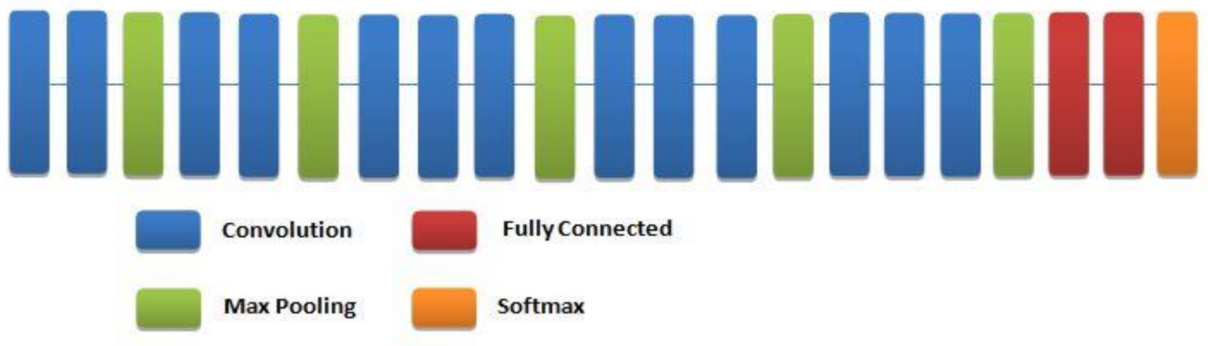

Fig. 1VGG16 Architecture

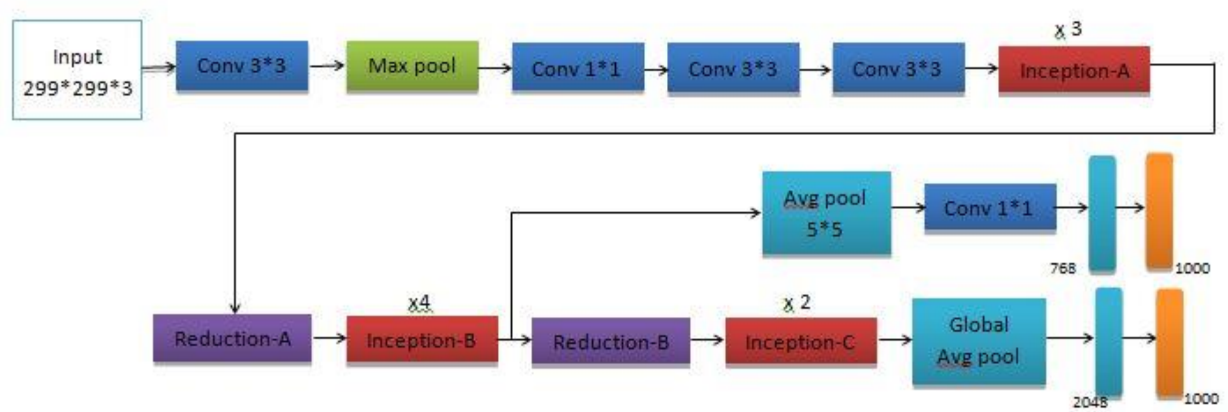

Fig. 2 Inception V3 architecture

Especially, Deep learning techniques can solve the problems of image classification with acceptable accuracy. These techniques are contrary to the traditional feature based supervised learning approaches like support vector machine, random forests etc. and are said to be end to end approaches as they do not need handcrafted feature as input. There exist many pre-trained deep learning architectures that have been designed, trained and tested on massive datasets for various image classification problems. These architectures consist of various layers through which representation of image content and its analysis is carried out. The Figure 1 of VGG16 architecture gives a glimpse of layers and their interconnection. The Figure 2 represents Inception V3 architecture, which is one of the popular Deep Learning architectures.

2. Related Work:

In the past few years, researchers were working on technological solutions for addressing various agricultural problems ranging from satellite image based disease analysis to smart irrigation. Digitization paved way for AI technology evolution, which intern offering solutions in every spectrum of life. Mohanty *et al*., [9] in their research, used two different CNN architectures i.e. AlexNet and GoogleNet for the identification of 26 different plant diseases. Baker and capel[11] the authors designed modelsfor group of crops such as (Cotton, Soybeans, Rice, Corn and Vegetables and Ground Fruits)using the USA and Switzerland-produced PlantVillage [10] region dataset. Due to the different leaf forms, variations, and climatic conditions, potato diseases vary by region. In [6] authors purposed a novel plant leaf disease identification model based on deep CNN to differentiate between healthy and unhealthy plants with 39 different classes of plant leaves and background images. Kamal *et al*., [7] developed models named Modified MobileNet and Reduced MobileNet for the identification of plant leaf diseases using convolution layers by modifying the MobileNet[12].Many Researchers have worked on potato crop diseases and built models using a particular dataset of PlantVillage. Sanjeev *et al*., [13] proposed a Feed-Forward Neural Network (FFNN) approach in the identification of two diseases early blight, late blight along with healthy leaves. The proposed model was trained and tested on an open access PlantVillage dataset. Tiwari *et al*., [14] used a transfer learning approach to extract the featuresand multiple classifiers of neural network for classification. The model also trained on the PlantVillage dataset to detect the early blight and late blight disease of potato plant. Barman *et al*., [15] developed self-build CNN (SBCNN) model to classify three different classes' early blight, late blight and healthy potato leaf diseases. To train model the PlantVillage dataset was used, however they did not validate their model on unseen data. Lee *et al*., [16] classified potato leaf disease from PlantVillage dataset using CNN. Rozaqi and Sunyoto[17] proposed CNN model and trained on a PlantVillage dataset to identify the diseases of potato and healthy class. Rashid *et al.*, [18] used a YOLOv5 image segmentation technique to extract, the potato leaves from the potato plant image. They developed novel deep learning technique using CNN to detect the diseases from potato leaf images. The potato leaf dataset collected from the Central Punjab Region of Pakistan. Islam *et al*., [19] developed a segment and multi SVM-based model to detect potato diseases. Khalifa [20] *et al*., proposed a CNN model to detect early blight and late blight along with a healthy class of leaves over the PlantVillage dataset.

Transfer learning is the process of utilizing the weights of a model that has already been trained but may not have used the same dataset. Many pre-trained deep learning architectures, including VGG16, VGG19, RESNET 50, MobileNet V2, InceptionV3 and others, are available for a popular image classification problem. Deep learning models classification accuracy is affected by the size of dataset, the variety of the data. The degree of class imbalance [21] analyzes the impact of the dataset size and it's variety on a model performance. Saleem *et al*., [22] presents a broad explanation of DL models uses to depict various plant diseases. Yang X and Sun M.[23] presented survey on deep learning approaches for different crop plantings.

3. Proposed Model for Potato Disease Classification

Dataset:

Performance of the deep learning models heavily depends upon an appropriate and valid dataset. The proposed model's performance is accessed using potato leaf images of a publicly available dataset called PlantVillage [10]. It was developed by Penn State University (US) and EPFL (Switzerland), which is a non-profit project. The dataset consist of JPG color images with 256 x 256 dimensions. It has total 38 classes of diseased and healthy leaves of 14 plants. Focus of this research is on the potato crop only. Potato leaves of 3 different classes (healthy, early blight, late blight) were considered for classification problem as shown figure 3.The sample size description of the dataset has been given in Table 1.

Table 1

Summary of the sample size of the dataset.

Sr. No.	Sources	Class Label	Samples
1.	PlantVillage	Early Blight	1000
		Late Blight	1000
		Healthy	152
	Total Samples		**2152**

(a) (b) (c)

Fig. 3 Samples collected from PlantVillage dataset (a) Healthy, (b) Early Blight and (c) Late Blight

Architecture:

There are various deep learning architectures that can classify images as well as make sense of them; Fang and Ramasamy [24] provide a comprehensive analysis of plant disease detection approaches with deep learning. In this paper noveldeep learning architecture had been proposed assimilating two of the popular architectures, namely VGG16 and Inception V3. These architectures were considered as they have been primarily used for image recognition. However, with the help of transfer learning and parameter optimization, they are adequate at other tasks, such as video classification, semantic segmentation, recognition of interest extraction, image indexing and retrieval, etc.The model is also compared with VGG16 and Inception V3 over similar dataset for analysis. The process involves various stages right from data partitioning to image classification as shown in figure 4.

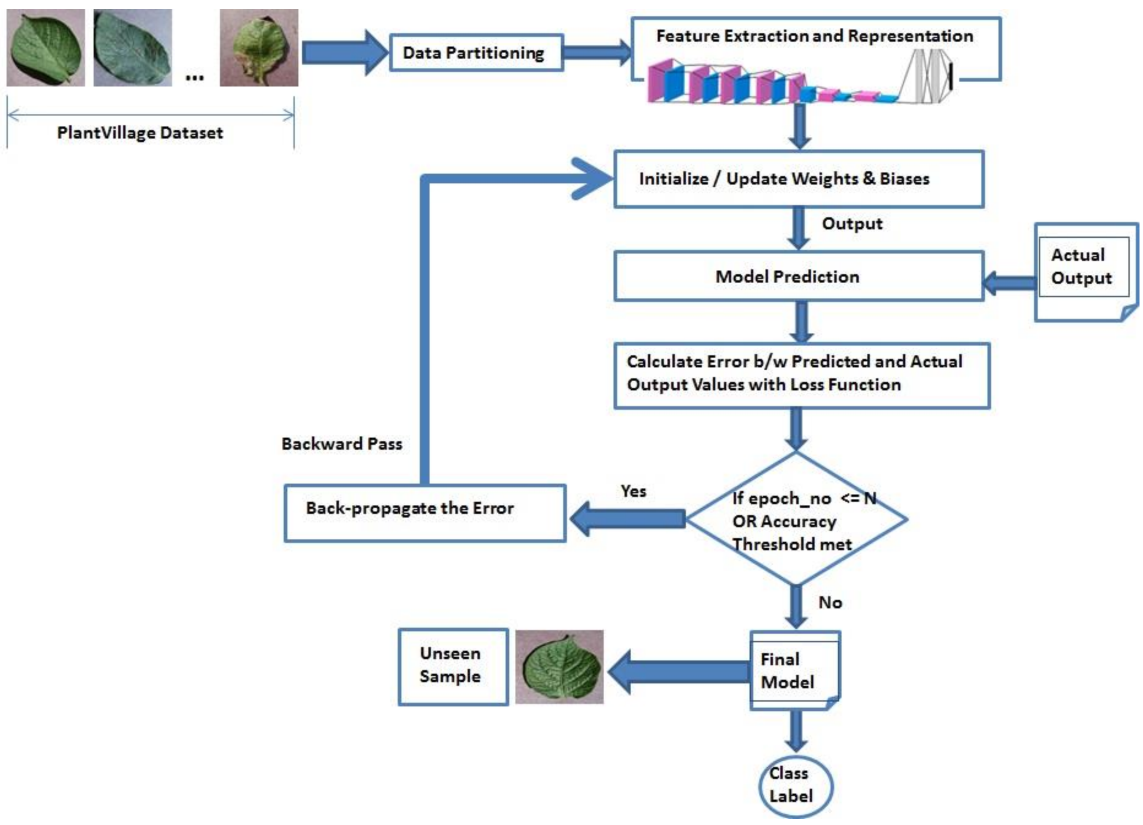

Fig. 4 Flow Diagram of Potato Leaf Disease Classification

In the backward pass, the weights (or parameters) in the model are adjusted to back-propagate the error that is computed using the predicted value and the expected value. The training process is repeated over a finite number of iterations, called epochs. Once training stops, the model's, accuracy is noted and it is exported to the disk.

The figure 5 representing proposed architecture with combination of different layers and their respective sequence. Convolutional neural networks can solve image classification challenges effectively when convolution, pooling, and regularization operations are conducted all over the signal processing pipeline.

This proposed model has been trained and tested on 2152 images of potato leaves taken from open access PlantVillage dataset for identification of two diseases early blight and late blight along with the healthy leaves. The Image Data Generator method of the Keras library in Python was used to apply several data augmentation strategies to the training set in order to reduce overfitting and increase the dataset's variety.

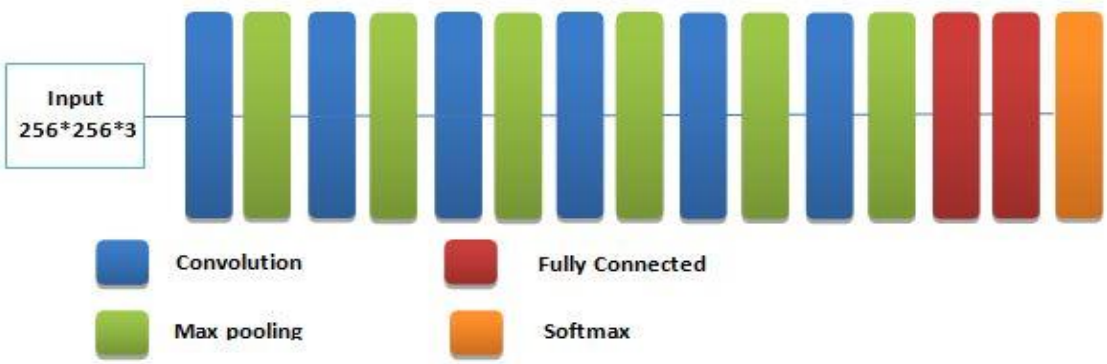

Fig. 5 The Proposed Deep Learning Model

For training, it inputs 256 x 256 pixel RGB image. The model has convolution layers having a kernel size of 3 x 3. Max-pooling is applied to reduce the size of the training vector space. The proposed model contains 6 convolution layers, 6 max-pooling and 3 fully connected layers as depicted in fig.5. The activation functions used in the convolutional and dense layers are ReLU (Rectified Linear Unit) and default learning rate i.e. 0.001. There are two fully connected layers, which are then followed by a SoftMax classifier and Dropout (0.2). The Convolution stride used is of one pixel. The proposed architecture learns around 183,747 parameters. The layers with trainable weights are convolution layers and then fully connected layers. These convolution layers and pooling layers used to extract the features and they will be followed by the dense layers for learning and prediction later. Softmax function is used to make final decision. The output produced by the previous layer is the input to the next layer in the sequence. Therefore nodes in each consecutive layer can recognize more complex and detailed features.

Algorithm:

PlantVillage dataset has 38 classes of diseased and healthy leaves of 14 plants. Focus of this research is on the potato crop only. Therefore, 1000 leaves for early blight, 1000 leaves for late blight, and 152 images of healthy leaves were selected for experimental purposes, as shown in Table 1.

Dataset = {Potato_Early_bligtht, Potato_Late_blight, Potato_Healthy}

Step 1: Load and Data Preprocessing

- Load dataset and resize them into 256 x 256 dimensions
- Split the dataset into train and test
- After standardization convert them into numerical vectors
- Apply data augmentation to the training set using ImageDataGenerator

Step 2: Generate Model

- Load all the layers in the model and other hyper-parameters
- Apply ReLU to avoid linearity and pooling for dimensionality reduction
- Use Dropout to reduce overfitting

Step 3: Model Compilation and Training

- Compile the model with Adam as the optimizer and sparse categorical cross-entropy as the loss function for training
- Model modifies weights via SGD after learning from data
- Each epoch records loss and accuracy

Step 4: Label Prediction

- Load the trained model
- Preprocess an image after loading it
- Predict the image's label

4. Experimental Results and Discussion

Experimental setup and tools

Proposed model has been implemented in Google colab using Keras and Tensorflow as a backend. Experiments were carried out on a machine having an AMD E2-6110 APU with AMD Radeon R2 Graphics, 16 GB RAM of NVIDIA and Windows 8 operating system. The models were trained up to 25 epochs and for the optimization, Adam optimizer is used. The entire raining was completed around 6 hours.

The prediction results for potato Early Blight, Late Blight and Healthy have been presented in Fig. 6, for each class of prediction, 4 images have been sampled out. It is evident from the given results that the proposed PoLeD CNN model accurately finds the correct labels on these images with the polling strategy.

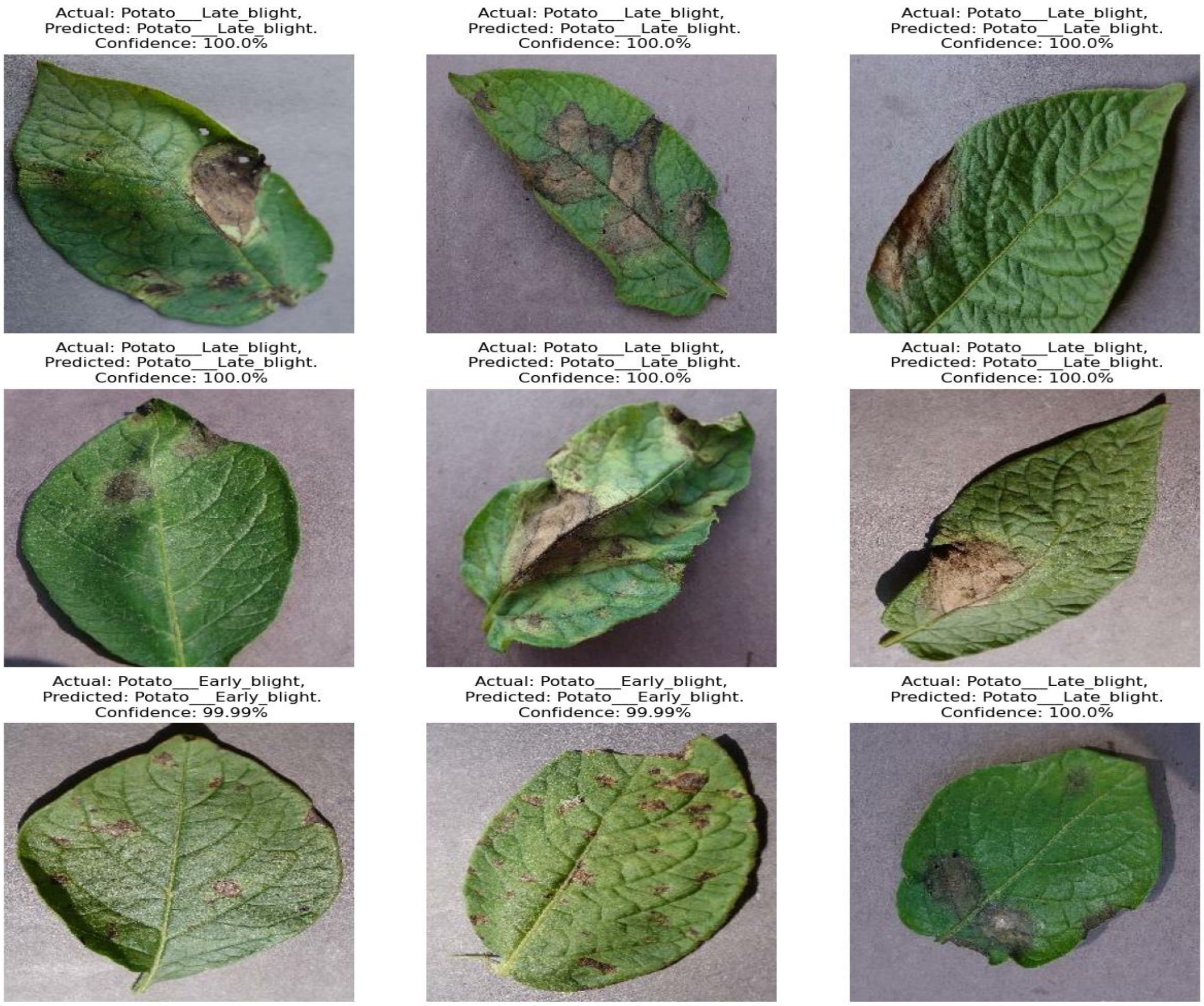

Fig. 6 Prediction result for Potato leaves of Early Blight, Potato Late Blight and Healthy classes.

Table 2 shows the various performance measures for the proposed PoLeD CNN model. It includes precision, recall and F1-score. F1-Score combines the precision and recall score of model. It transmits the balance between precision and recall.

$$\text{F1-Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Table 2

	Performance Measures	Early Blight	Late Blight	Healthy	Average
With Data Augmentation	Accuracy	99.49%	100%	100%	99.61%
	Precision	99%	100%	100%	-
	Recall	100%	100%	99%	-
	F1-Score	100%	99%	100%	-
Without Data Augmentation	Accuracy	92.87%	85.47%	93.25%	91.23%
	Precision	88%	94%	90%	-
	Recall	94%	85%	89%	-
	F1-Score	95%	92%	94%	-

Classification accuracies, precision, recall and F1-Score of the proposed PoLeD CNN model

Comparative Analysis:

With the similar setup, having same training dataset, the performance of VGG16 and Inception V3 architectures have been studies and compared with proposed model. The results of the same are presented in Table 3.

Table 3

Architectural comparison of three different models.

Sr. No.	Model	Accuracy Rate	Total Params	Training Params	Non- Trainable Params
1.	VGG16	79.52	14,789,955	75,267	14,714,688
2.	Inception V3	64.25	21,858,083	55,299	21,802,784
3.	Proposed CNN	99.61	183,747	183,747	0

Comparison with other state-of-art methods

The proposed model performance was compared with existing techniques from the literature over potato leaf disease detection. It is observed that the proposed deep learning model performed significantly well as compared to state-of-the-art techniques and it achieved an accuracy 99.83% compared to existing studies as shown in Table 8. The proposed CNN model dominated the existing techniques, thus achieving 99.83% accuracy with fewer parameters, i.e., 183,747, leading to a lower computational cost and the highest accuracy compared to existing models.

Table 4

Comparison with state-of-the-art techniques

Existing Study	Total Parameters	Accuracy
Sanjeev*et al.*,[13]	-	96.50%
Tiwari*et al.*,[14]	143,667,240	97.80%
Khalifa*et al.*,[20]	-	98.00%
Lee *et al.*,[16]	10,089,219	99.00%
Rashid*et al.*,[18]	8,578,611	99.75%
PoLeD CNN	183,747	99.61%

5. Conclusion

The research paper presents a fast and straightforward novel deep convolutional architecture for potato leaf disease classification over the dataset taken from PlantVillage dataset. The proposed method is used to classify 3 potato disease classes (Early Blight, Late Blight and Healthy). The advantage of the proposed model is that it can optimally learn from the huge training data with moderate usage of resources and with data augmentation technique, thus accuracy of disease recognition is 99.61% with high precision, recall, F1-score. It has a minimal number of parameters and was simpler than the state-of-the-art models, saving a substantial computational cost and speed. Therefore the proposed model can be used as a good tool for potato leaf disease classification.

References

1. Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., urRehman, M. H., &Javed, K. (2018, October 1). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Computers and Electronics in Agriculture*. Elsevier B.V. https://doi.org/10.1016/j.compag.2018.07.032
2. Oerke, E.C. Crop losses to pests. J. Agric. Sci. 2006, 144, 31–43.
3. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. Inf. Process. Agric. 2018, 5, 354–371.
4. Chouhan, S.S.; Kaul, A.; Singh, U.P.; Jain, S. Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. IEEE Access 2018, 6, 8852–8863.
5. Sasaki, Y.; Okamoto, T.; Imou, K.; Torii, T. Automatic diagnosis of plant disease-Spectral reflectance of healthy and diseased leaves. IFAC Proc. Vol. 1998, 31, 145–150.
6. Geetharamani, G.; Pandian, A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Comput. Electr. Eng. 2019, 76, 323–338.
7. Kamal, K.; Yin, Z.; Wu, M.; Wu, Z. Depthwise separable convolution architectures for plant disease classification. Comput. Electron. Agric. 2019, 165, 104948.
8. Khamparia, A.; Saini, G.; Gupta, D.; Khanna, A.; Tiwari, S.; de Albuquerque, V.H.C. Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network. Circuits Syst. Signal Process. 2019, 39, 818–836.
9. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. Front. Plant Sci. 2016, 7, 1419.
10. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv 2015, arXiv:1511.08060.
11. Baker, N.T.; Capel, P.D. Environmental Factors that Influence the Location of Crop Agriculture in the Conterminous United States; US Department of the Interior, US Geological Survey: Reston, VA, USA, 2011.
12. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv 2017,

arXiv:1704.04861.

13. Sanjeev, K.; Gupta, N.K.; Jeberson, W.; Paswan, S.Early Prediction of Potato Leaf Diseases Using ANN Classifier.*Orient. J.Comput.Sci.Technol.***2020**,*13*,2–4.
14. Tiwari, D.; Ashish, M.; Gangwar, N.; Sharma, A.; Patel, S.; Bhardwaj, S. Potato leaf diseases detection using deep learning.InProceedingsofthe2020 4th InternationalConferenceonIntelligentComputingandControlSystems(ICICCS),Madurai,India,13-15May2020;pp.461–466.
15. Barman, U.; Sahu, D.; Barman, G.G.; Das, J. Comparative Assessment of Deep Learning to Detect the Leaf Diseases of PotatobasedonDataAugmentation.InProceedingsofthe2020InternationalConferenceonComputationalPerformanceEvaluation(ComPE),Shillong,India,2–4July2020;pp.682–687.
16. Lee, T.Y.; Yu, J.Y.; Chang, Y.C.; Yang, J.M. Health Detection for Potato Leaf with Convolutional Neural Network. In Proceedingsof the 2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), Rajpura,India,7–15February2020;pp.289–293.
17. Rozaqi, A.J.; Sunyoto, A. Identification of Disease in Potato Leaves Using Convolutional Neural Network (CNN) Algorithm. InProceedingsofthe2020 3rd InternationalConferenceonInformationandCommunicationsTechnology(ICOIACT),Yogyakarta,Indonesia,24–25November2020;pp.72–76.
18. Rashid, J., Khan, I., Ali, G., Almotiri, S. H., Alghamdi, M. A., &Masood, K. (2021). Multi-level deep learning model for potato leaf disease recognition. *Electronics (Switzerland)*, *10*(17). https://doi.org/10.3390/electronics10172064
19. Islam, M.; Dinh, A.; Wahid, K.; Bhowmik, P.Detection of potato diseases using image segmentation and multiclass supportvector machine. In Proceedings of the 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE),Windsor,ON,Canada,30April–3May2017;pp.1–4.
20. Khalifa, N.E.M.; Taha, M.H.N.; Abou El-Maged, L.M.; Hassanien, A.E. Artificial Intelligence in Potato Leaf Disease Classification:A Deep Learning Approach. In *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*; Springer:Berlin/Heidelberg,Germany,2021;pp.63–79.

21. Barbedo JGA. Impact of dataset size and variety on the effectiveness of deeplearning and transfer learning for plant disease classification. ComputElectronAgric 2018;153:46–53.
22. Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant Disease Detection and Classification by Deep Learning. *Plants* **2019**, *8*, 468. https://doi.org/10.3390/plants8110468
23. Yang X, Sun M. A survey on deep learning in crop planting. In: IOP conferenceseries: materials science and engineering, 490. IOP Publishing; 2019, April. p. 6.062053.
24. Fang, Y.; Ramasamy, R.P. Current and Prospective Methods for Plant Disease Detection. *Biosensors* **2015**, *5*, 537-561. https://doi.org/10.3390/bios5030537ImageNet.
25. ImageNet.http://www.image-net.org