J. Phytol. Res. **34** (2): 177- 185, 2021 **ISSN 0970-5767**

DEEP CONVOLUTION NEURAL NETWORK-BASED PREDICTION AND CLASSIFICATION OF BLIGHT DISEASE OF POTATO

JYOTI LAKHANI[1] and DHARMESH HARWANI[2]*

[1]Department of Computer Science, Maharaja Ganga Singh University, Bikaner
[2]Department of Microbiology, Maharaja Ganga Singh University, Bikaner
*Corresponding Author's E mail: dharmesh@mgsubikaner.ac.in

The blight disease of potato is one of the major threats to food security, the rapid identification of which remains difficult in many resource-limited areas. In the present paper, a deep learning-enabled, object detection model for a multi-class blight disease of potato has been proposed. Attempts have been made to improve the mobile-assisted *in situ* detection of the disease. The proposed model employs a state-of-the-art computer vision algorithm. A public dataset of 2,152 images from healthy and infected leaves was used to train a deep convolution neural network model to predict the health state of a potato plant. The publically available benchmark dataset was used to train the deep learning model to detect the early blight and late blight disease of the potato plant. When tested, the proposed model was observed to best fit, among others, for a 60-20-20 train-test-validation split with 99.31% accuracy and 99.9% confidence. The proposed end-to-end approach model may be employed in the future as an effective and efficient android application that can assist farmers to detect and classify various diseases in crop plants rapidly.

Keywords: Artificial intelligence, Blight disease of potato, Convolution neural networks, Deep learning.

Introduction

Modern technologies have given human society the ability to produce enough food to meet present and future demands. However, food security remains threatened by several factors including climate change[1], declining pollination[2], plant diseases[3], and others. Many efforts have been made to prevent crop loss due to various plant diseases. Therefore, the identification of a plant disease correctly, when it first appears, is a crucial step for its efficient management. There are tremendous advancements in the field of computer vision, and object recognition that can fulfill this task. In the present communication, we have used a deep learning approach that utilizes 2,152 images of healthy and diseased potato plants for early detection of the disease. The images for early and late blight disease are made openly available through the project PlantVillage[4]. An example of a healthy, early, and late blight disease of the potato plant is shown in Fig. 1. A farmer needs to identify the disease at its early stage (early blight), otherwise it will be difficult to treat the plant at the later stages of the disease (late blight). Previously, convolution neural networks and deep learning approaches such as CNN, Yolo, and Xgboost have been already used to detect the blight disease of potato using image segmentation techniques. For a good end-to-end disease detection system, where the end-user is a poor farmer, it is important to keep the method as simplest as possible. To fulfill this task, we have kept the architecture of the proposed neural network model simpler and tried to optimize the performance without using complex algorithms. Using a combination of specialized activation functions for optimization the trained model achieves an accuracy of 100% for 75-15-10 split of the dataset for training-testing-validation. However, we cannot deny the overfitting issue. Consequently, 10 different training-testing-validation spits were also trained in the present work. Surprisingly, most of the time, 100% accuracy was observed in either testing or validation dataset. To ignore any possibility of over fitting, a 60-20-20 split dataset was finally considered that exhibited 99.85% training, 99.26% validation, and 99.31% testing accuracy. In this way, an end-to-end learning approach of deep neural networks was implemented in the present empirical study to identify and detect early and late blight disease of potato.

Material and Methods

Dataset description:

A total of 2152 images of healthy and diseased (early and late blight) potato plant leaves were analyzed. A representation of these images has been presented in Fig. 1. The potato disease dataset is a subset of the PlantVillage dataset freely available at Kaggle. It is a computer vision dataset consisting of images of potato plant leaves having 'Healthy', 'Early-Blight' and 'Late-Blight' categories. The 'Healthy', 'Early-Blight' and 'Late-Blight' categories consist of 152, 1000, and 1000 numbers of images of healthy leaves and leaves exhibiting early blight and late blight respectively. The images are resized to 256 x 256 pixels to support the proposed optimized deep neural network.

Experimental Setup:

In the proposed deep learning model, a convolution neural network is designed with an input layer, an output layer, and many intermediate hidden layers. Each node in the input layer accepts an image of the leaf from the diseased plant. The nodes in the middle layers are mathematical functions that perform the mapping between the input image of the leaf from the diseased plant and it is classified as a disease label. The neural network parameters are tuned and optimized for performance during the training process. Selecting correct parameters to optimize the performance of the neural networks is a challenging task[5,6]. For an accurate image classifier to accurately predict plant disease, it is imperative to train the neural network using a large dataset of diseased and healthy plants. To fulfill this requirement, the Plant Village project was started to collect thousands of images of healthy and diseased crop plants and make them openly available[4]. In the present study, the classification of early and late blight disease of potato was performed using 2,152 plant images as an input to convolution neural-network based deep learning models. The performance of the model to predict the healthy or diseased state of the plant and its accurate identification is measured in the form of accuracy rate.

Proposed Deep Learning Model for Plant Disease Classification:

The experimental work has been performed on Google's Colab[7] research platform. Keras[8] framework was used which is available as a part of Tensorflow[9]. Python programming was used for all the experiment work. The dataset used in the work was split into three partitions- training, validation, and testing wherein the training split of the dataset were used to train the neural network model. All dataset items in the training set were given to the model as an input with its class label. During training, the neural network analyzes the dataset by associating the image with its class label. After the training session, validation of the model was performed. In the validation step, the model was fed with some known images to predict the class label. In the testing step, the proposed model was given a subset of the dataset which was unknown to the neural network. In all three stages, the accuracy of the detection of the model was calculated; consequently, these were called training, validation, and testing accuracy. To train and test the proposed model following train-test splits of the dataset was performed:

i) **60-30-10 train-test split**- in which 60% of the whole dataset was used for training, 30% for validation, and 10% for testing

ii) **60-20-20 train-test split**- in which 60% of the whole dataset was used for training, 20% for validation, and 20% for testing

iii) **65-20-15 train-test split**- in which 65% of the whole dataset was used for training, 20% for validation, and 15% for testing

iv) **65-15-20 train-test split**- in which 65% of the whole dataset was used for training, 15% for validation, and 20% for testing

v) **70-10-20 train-test split**- in which 70% of the whole dataset was used for training, 10% for validation, and 20% for testing

vi) **70-20-10 train-test split**- in which 70% of the whole dataset was used for training, 20% for validation, and 10% for testing

vii) **75-15-10 train-test split**- in which 75% of the whole dataset was used for training, 15% for validation, and 10% for testing

viii) **75-10-15 train-test split**- in which 75% of the whole dataset was used for training, 10% for validation, and 15% for testing

ix) **80–10-10 train-test split**- in which 80% of the whole dataset was used for training, 10% for validation, and 10% for testing

x) **80–15-5 train-test split**- in which 80% of the whole dataset was used for training, 15% for validation, and 5% for testing

Before the train-test partitioning, the dataset was preprocessed by fixing the image size to 256 x 256 pixels. The dataset was partitioned in such a way that during partition the shuffling of data items could be performed with the shuffle_size parameter set to 100. The batch size for the experiment was set to 32 with 3 channels and 50 epochs wherein each epoch runs 54 times. During preprocessing, resizing (256 x 256) and reshaping (1.0/255) was also performed. Further, data augmentation was performed in which horizontal and vertical random flip and random rotation at scale 0.2 were implemented.

The input layer in the sequential model was a convolution 2d layer with 32 filters, a kernel of shape 3 x 3 with a 'ReLU' activation function. In the second and third layers, 64 filters were used with a kernel of shape 3 x 3 with a 'SeLU' activation function. A two-dimensional MaxPooling was performed in between every two convolution neural network layers. In the fourth and fifth conv2d layers, the filter size was reduced to 32 with activation function 'Sigmoid'. The output end consisted of two dense layers with activation functions 'Tanh' and 'Softmax' respectively. The summary of the proposed deep learning model has been shown in table 1. The layered diagram of the proposed model has been also given in fig. 2.

The 'ReLU' activation function in the input layer later supports the classification between healthy and diseased leaves of the plant. The 'ReLU' functions are good at classification problems that do not trigger all neurons at the same time, thus considered computationally efficient than the other activation functions. In the next two layers, 'SeLU' activation functions were used which are faster and easy to implement. The vanishing gradient points of 'ReLU' do not appear in the 'SeLU' activation function. In the next two layers, the 'Sigmoid' activation function was used which deals with non-linearity. Subsequently, the Tanh activation function was employed which removes the gradient direction issue if it appears due to the 'Sigmoid' activation function. In the last output layer, the 'Softmax' activation function, a specialized function for multi-class classifications to handle the three discrete classes in the data, was used. The

Table 1: The Input-output shape and parameters used in the layers of the proposed deep convolution neural network

Layer (type)	Input Shape	Output Shape	Param#
Sequential	(32, 256, 256, 3)	(32, 256, 256, 3)	0
Sequential	(32, 256, 256, 3)	(32, 256, 256, 3)	0
Conv2D	(32, 254, 254, 3)	(32, 256, 256, 32)	896
MaxPooling2D	(32, 254, 254, 3)	(32, 127, 127, 32)	0
Conv2D	(32, 127, 127, 32)	(32, 125, 125, 64)	18496
MaxPooling2D	(32, 125, 125, 64)	(32, 62, 62, 64)	0
Conv2D	(32, 62, 62, 64)	(32, 60, 60, 64)	36928
MaxPooling2D	(32, 60, 60, 64)	(32, 30, 30, 64)	0
Conv2D	(32, 30, 30, 64)	(32, 28, 28, 32)	18464
MaxPooling2D	(32, 28, 28, 32)	(32, 14, 14, 32)	0
Conv2D	(32, 14, 14, 32)	(32, 12, 12, 32)	9248
MaxPooling2D	(32, 12, 12, 32)	(32, 6, 6, 32)	0
Flatten	(32, 6, 6, 32)	(32, 1152)	0
Dense	(32, 1152)	(32, 64)	73792
Dense	(32, 64)	(32, 3)	195
Total params: 158,019 **Trainable params: 158,019** **Non-trainable params: 0**			

Healthy Leaf	Early Blight	Late Blight

Figure 1: Potato plant with a healthy leaf (left) and leaves with early (middle) and late blight (right)

arrangement of the various layers of the proposed deep convolution neural network to detect and classify blight disease of the potato plant has been shown in Fig 2. The Input-output shape and parameters used for the proposed model have been given in table1.

Results and discussion

The proposed model was fitted and tested for ten different train-test splits. The accuracy of the model was tracked at the end of each of the 50 epochs. The overall accuracy of the complete period of training is an average of accuracy achieved at the end of each epoch. The experiment results are shown in Table 2. For each epoch, we have calculated percent accuracy and percent loss. The model correctly predicted the blight disease of potato with 100 percent accuracy for 75-15-10 split for training, testing, and validation dataset in 50 epochs and 54 steps. The high accuracy of the split

Figure 2: Arrangement of the various layers in the proposed deep convolution neural network model to predict blight disease of a potato plant

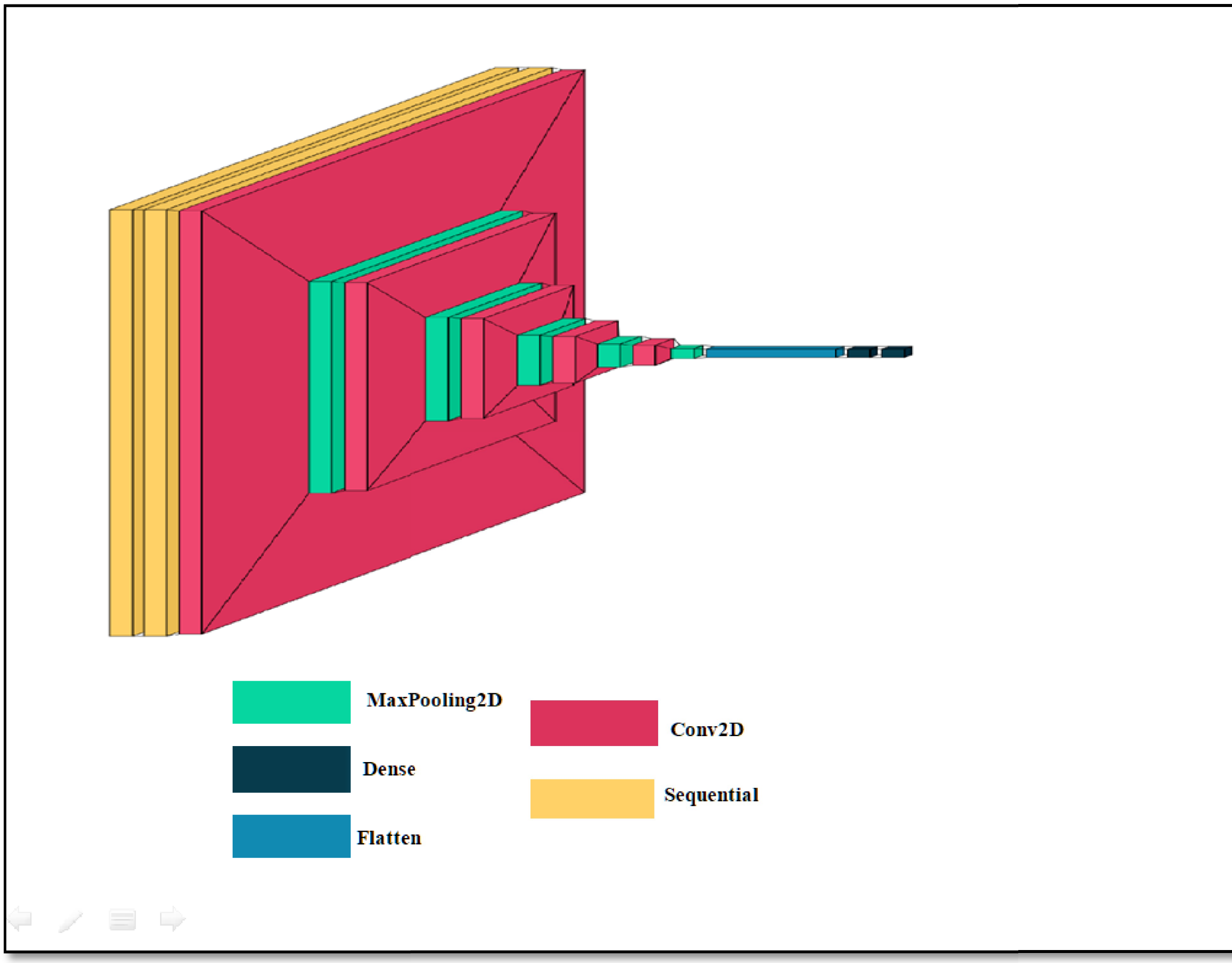

Table 2: Summary of results for various train-test-validation splits of the dataset

S.No.	Dataset			% Accuracy (Average)		
	Train	Test	Validation	Train	Test	Validation
1	60	30	10	99.85	99.40	100
2	**60**	**20**	**20**	**99.85**	**99.31**	**99.26**
3	65	20	15	95.72	94.53	95.31
4	65	15	20	99.71	99.61	100
5	70	10	20	99.83	100	100
6	70	20	10	99.59	100	98.96
7	75	10	15	99.88	100	100
8	75	15	10	100	100	100
9	80	10	10	98.61	97.50	98.21
10	80	15	5	98.88	100	100

may further be assigned to the overfitting issue. As given in Table 2, for 70-30, 75-25 and, 80-20 partitions; the validation accuracy is more than the testing accuracy which possibly depicted an over fitting issue. Therefore, the results of 70-30, 75-25, and 80-20 train-test split models were ignored. Moreover, for 65-25-10 and 65-15-20 split models over fitting issue cannot be denied.

Given these results, only two models 60-20-20 and 65-20-15 train-test-validation split can be considered as the best fit. By viewing the accuracy rates of each split, the model with a 60-20-20 train-test-validation split can be considered as the best fit (highlighted in Table 2). The training and validation accuracy as well as training and validation loss has been plotted against the epochs (Fig. 3). The 60-20-20 train-test-validation split model predicted the test dataset with 99.31 percent accuracy and 99.9 percent average confidence. The result as the actual and predicted cases of the test data with percent confidence is shown in fig. 4.

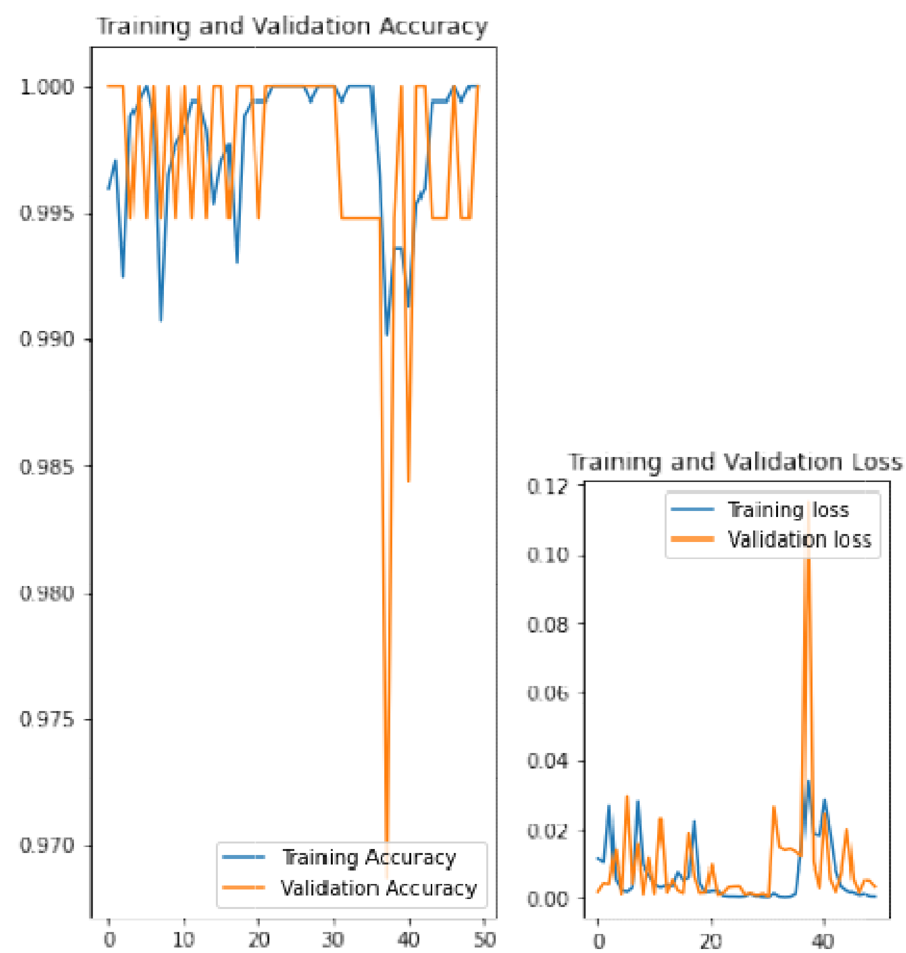

Figure 3: Training and validation accuracy and training and validation loss have been plotted against epochs. Training accuracy shows the performance in terms of accuracy of the proposed model with known dataset. The validation accuracy is the performance of the proposed model with the unknown data. The training loss reflects the performance loss of the proposed model during the training session. The validation loss of the proposed model demonstrates the performance loss of the model during the validation session.

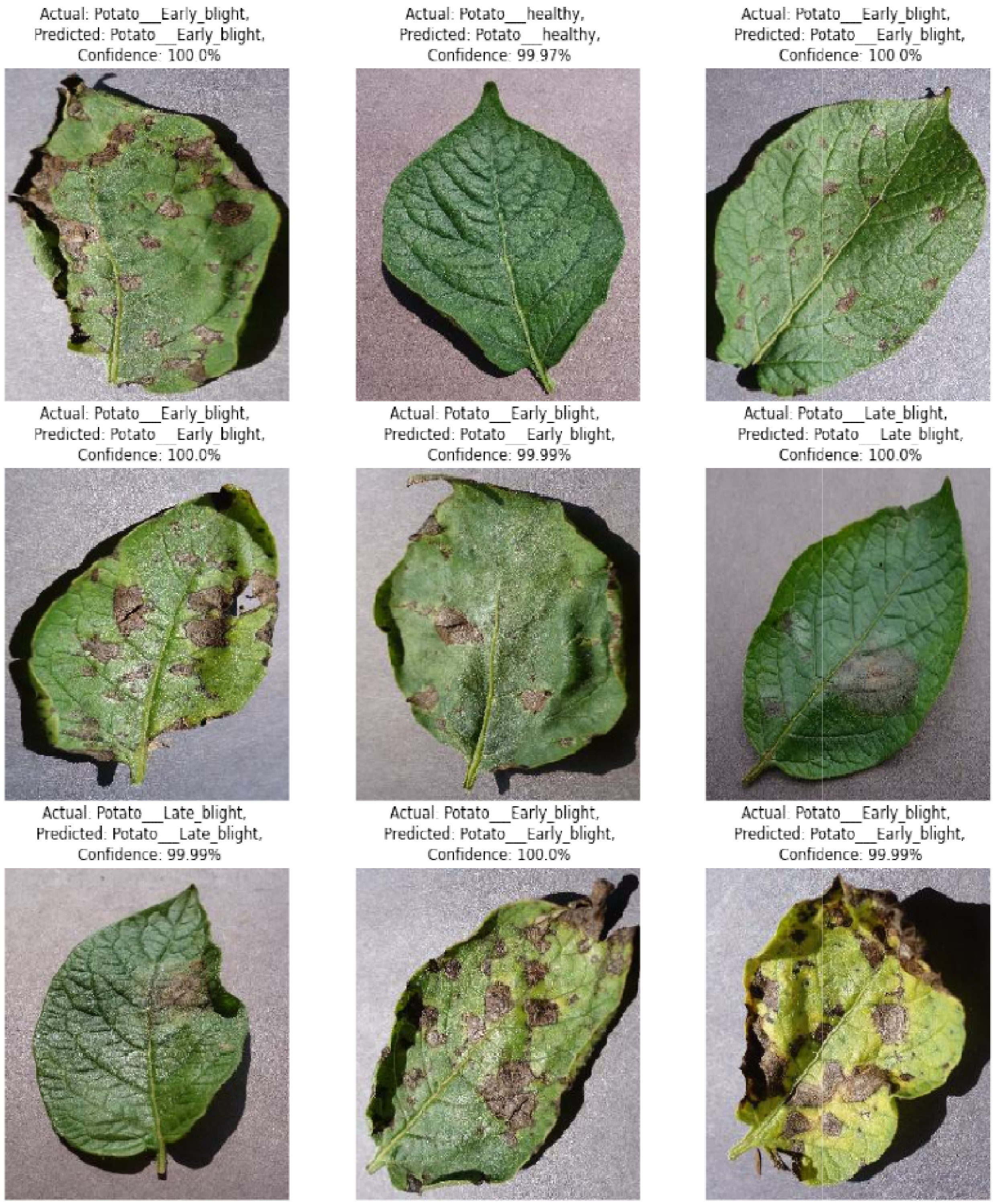

Figure 4: The actual and predicted cases of the test data with percent confidence

Previously, many researchers have performed potato crop disease classification using convolution neural networks. The observed training and testing accuracy of their proposed model was 99.47% and 99.8% respectively[10]. The convolution neural network model with the YOLOv5 image segmentation technique was also used which showed 99.75% accuracy[11]. While shallow CNN with Xgboost model showed a 98.74% accuracy rate[12]. Other workers have also performed enhanced field-based detection of the disease with complex backgrounds[13]. For other related research work, please consider[14-19]. Interestingly, not only blight disease of the potato plant but prediction and classification of other diseases in tomato[16,20-22], mango[23], cotton[24], bell pepper[16], maize[25] plants have also been addressed.

Conclusion

It is important to accurately detect and classify the disease type in the crop plants for its proper management. The late identification of diseases may cause less productivity which may affect directly the wealth of farmers and indirectly the economy of our country. The technological advancements in the area of computer vision and neural networks have enormously improved the methods in the field of disease identification and detection in plants. The present paper also extends the importance of convolution neural networks and deep learning method to accurately predict early and late blight disease of potato plants. The proposed model was observed to be the best fit for a 60-20-20 train-test-validation split with 99.31% accuracy and 99.9% confidence. The present research work can be extended in the future to design an end-to-end android application for farmers to detect and classify disease in crop plants by simply clicking and uploading an image. The work can be extended further as well to work with gray-scale using segmented image datasets to improve its performance.

References

1. Tai A P, Martin M V and Heald C L 2014, Threat to future global food security from climate change and ozone air pollution. Nat. Clim. Chang. 4 817–821.
2. Report of the Plenary of the Intergovernmental Science-PolicyPlatform on Biodiversity Ecosystem Services on the work of its fourth session (2016). In: Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Fourth session. Kuala Lumpur.
3. Strange R N and Scott P R 2005, Plant disease: a threat to global food security. Phytopathology 43 83–116.
4. 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
5. LeCun Y, Bengio Y and Hinton G 2015, Deep learning. Nature 521 436–444.
6. Alzubaidi, Laith, et al. 2021, Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. big Data 8.1 1-74.
7. Google Colab [https://colab.research.google.com/?hl=no-NO#scrollTo=GJBs_flRovLc]
8. GitHub - keras-team/keras: Deep Learning for humans [https://colab.research.google.com/?hl=no-NO#scrollTo=GJBs_flRovLc]
9. Martín et al. 2015, TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
10. Agarwal et al. 2020, Potato crop disease classification using convolutional neural network. In: Smart Systems and IoT:

Innovations in Computing. Springer, Singapore, pp 391-400.
11. Rashid et al. 2021, Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. Electronics 10.17 2064.
12. Hassan et al. 2021, Plant Disease Identification Using Shallow Convolutional Neural Network. Agronomy 11.12 2388.
13. Johnson et al. 2021, Enhanced Field-Based Detection of Potato Blight in Complex Backgrounds Using Deep Learning. Plant Phenomics 2021.
14. Kukreja et al. 2021, Potato Blight: Deep Learning Model for Binary and Multi-Classification. 8th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2021.
15. Barman et al. 2020, Comparative Assessment of Deep Learning to Detect the Leaf Diseases of Potato based on Data Augmentation. 2020 International Conference on Computational Performance Evaluation (ComPE). IEEE.
16. Ravichandran R R, Norzihani Y and Nuzulha K I 2021, Detection using Convolution Neural Network for Plant Disease Identification in Tomato, Potato and Bell Pepper. In: Proceedings of the 3rd International Conference on Intelligent and Interactive Computing 2021 (UTeM Press). UTeM Press, 2021.
17. Khamparia et al. 2020, Seasonal crops disease prediction and classification using deep convolutional encoder network. Circuits, Syst. Signal Process. 39.2 818-836.
18. Labhsetwar et al. 2021, Performance Analysis of Optimizers for Plant Disease Classification with Convolutional Neural Networks. 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE). IEEE, 2021.
19. Asif, Md K R, Asfaqur M R and Most H H 2020, CNN based Disease Detection Approach on Potato Leaves. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE.
20. Agarwal, et al. 2020, ToLeD: Tomato leaf disease detection using convolution neural network. Procedia Comput. 167 293-301.
21. Agarwal M, Gupta S K and Biswas K K 2020, Development of Efficient CNN model for Tomato crop disease identification. Sustain Comput-Infor 28 100407.
22. Brahimi M, Kamel B and Abdelouahab M 2017, Deep learning for tomato diseases: classification and symptoms visualization. Appl. Artif. Intell. 31.4 299-315.
23. Singh et al. 2019, Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease. IEEE Access 7 43721-43729.
24. Deeba K and Amutha B 2020, ResNet-deep neural network architecture for leaf disease classification. Microprocess. 103364.
25. DeChant et al. 2017, Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathology 107.11 1426-1432.