Section: Research Paper
ISSN 2063-5346

Early Recognition and Detection of Potato Leaf Disease Using Machine Learning

Dr. Chandra Prakash Patidar, Dr. Ashish Kumar Jain, Pritee Sawle
Information Technology, Computer Engineering, Computer Engineering
Institute of Engineering & Technology,
Devi Ahilya University,
Indore
cpatidar@ietdavv.edu.in, ajain@ietdavv.edu.in , sawlepritee123@gmail.com

DOI: 10.48047/ecb/2023.12.si4.1604

Abstract - Potatoes are the world's famous vegetable. A significant issue for potatoes is the identification of potato leaf disease. The leaves of potato plants display the signs of several illnesses, such as early blight, late blight, and others. Early detection of these outbreaks and quick action can prevent the farmer from suffering significant financial losses. Our proposed model would employ image processing techniques to accurately identify and categories potato leaf diseases. This study uses the CNN (Convolutional Neural Network) model, which is used for image classification and performs better than other approaches in terms of accuracy, to detect disease in images of potato leaves. This model separates potato leaf characteristics into normal and diseased ones. The potato plant leaf is classified as either normal or unhealthy after the images are evaluated using the provided algorithm. This experiment has achieved an average accuracy of 98%.

Keywords - Machine Learning, Image Processing, Convolutional Neural Network

I. INTRODUCTION

Agriculture has a significant impact on the economic growth of any nation. It is a subject that has a big impact on a nation's economy. A variety of factors affect the variety and volume of crops cultivated. Potatoes are well recognized around the world and are also an important fundamental meal in many nations. Potatoes are also known as the root of all veggies.The amount of demand is expanding globally on a daily basis, and it is also necessary to export as much as our region can, thus the important component is increasing potato production. However, due to major illnesses like late blight, early blight, and others, productivity and export levels have fallen in recent years. In that scenario, the level of output is severely impacted. Farmers are also suffering as a result of this. The affected potato leaf may occasionally show signs of the illness. Spots on the plant's leaves are still another possibility. Many diseases, including early and late blight, present in tiny oval, circular, and varied shapes. Early and late blight on potatoes are the two most typical kinds. Early blight symptoms are typically small, black patches, but later symptoms include blistering that looks like it has been burned by hot water and eventually decomposition and drying out. Farmers will greatly benefit from the deep learning model that will be offered to distinguish these diseases from potato leaves. The investigation requires a significant number of images because its primary focus is on images. The three categories of processed images are as follows. They are known as early, late, and healthy blight. The total quantity of images is divided into three segments: training, testing, and validation. About 80% of the photographs will be used for training, while the remaining 20% will be used for testing and validation. Our proposed model would distinguish between normal and sick potato leaves.

Our proposed model is divided into the following sections: In Section II, research that has recently been done to classify and detect plant diseases is discussed. In Section III, the suggested methodology and techniques are explained. The Experimental Setup is discussed in section IV.

II. LITERATURE REVIEW

Many techniques have now been developed in an effort to detect diseases in plant leaves.

Xin Li et al. [1] examined data sets of healthy leaves and the diseases apple grey-spot, black star, and cedar rust, to recognize and categorize apple leaf disorders. For analysis and advancement, models like ResNet, VGG (Visual Geometry Group) convolutional neural networks, SVM (Support Vector Machine) method classifier, and others were used. ResNet-18, the ResNet version used in the final test, had fewer layers and had better recognition results, with a 98.5% accuracy rate.

V. V. Srinidhi et al. [2] Two proposed deep convolutional neural network models, EfficientNet and DenseNet, are intended to accurately identify four distinct apple crop illnesses from photographs of an apple tree leaf. Among the categories are "healthy," "scab," "rust," and "Multiple diseases." Using image annotation and data augmentation techniques as Canny Edge Detection, Motion blur, and Rotating, the dataset for apple leaf disease in this study is improved. Based on a bigger dataset, EfficientNetB7 and DenseNet models are recommended because they provide performance of 99.8% and 99.75%, respectively, and overcome issues with convolutional neural networks.

K. Ahmed et al. [3] explain a system that uses machine learning to look for illness in rice leaves. Three of the most common diseases impacting rice plants are leaf smut, bacterial leaf blight, and brown spot disorders, according to this research. The data was composed of images of broken rice leaves against a white background. The dataset was used to train many machine learning algorithms, including KNN (K-Nearest Neighbor), J48 (Decision Tree), Naive Bayes, and Logistic Regression, after the required pre-processing. After 10-fold cross-validation, the decision tree approach produced results on the test dataset with an accuracy of above 97%.

S. V. Militante et al. [4] proposed a quick way to identify numerous illnesses in a variety of plant species. Apple, corn, grapes, potatoes, sugarcane, and tomatoes were specifically targeted for detection and recognition by the system. Numerous plant diseases are also detectable by the technique. With the use of 35,000 photos of disease-free and diseased plant leaves, the researchers were able to train deep learning models to identify plant diseases and their absence. The trained model has achieved an accuracy rate of 96.5% and was able to distinguish and identify the plant variety and the types of diseases the plant was affected by with up to 100% accuracy.

R. A. Sholihati et al. [5] used deep learning and the VGG16 and VGG19 convolutional neural network architectural model, based on the condition of the leaves, to categorize the four different illnesses that affect potato plants. The experiment's 91% average accuracy demonstrates the viability of the deep neural network strategy.

J. M. Al-Tuwaijari et al. [6], proposed system for classifying and identifying plant leaf diseases utilizing deep learning and image processing methods is presented. Two classification methods are included in the suggested system, and they are contrasted. The first approach used the SVM algorithm and had a number of steps before reaching the classification stage. The CNN (Convolution Neural Network) was utilized to be labeled in the second strategy. These two approaches categorized 15 separate categories, including 12 categories for plant illnesses that were found, including bacteria, fungi, and other pathogens, and three categories for healthy leaves.

M. K. R. Asif et al. [7], the suggested approach will properly identify and diagnose infections in potato leaf stands. Since CNN routinely beats other algorithms when it comes to image classification, it is employed in this work to identify the disease from photographs of the potato leaf. Five algorithms are used by us: AlexNet, VggNet, ResNet, LeNet, and the Sequential model we recommend. This model's accuracy was 97%, which is a decent level.

S. Chakraborty et al. [8] Describe a technique that makes use of image processing and machine learning to identify diseases in contaminated apple leaves. Apple leaves that are both sick and healthy can be differentiated using this method. The processing of the image, which includes the Otsu thresholding method and histogram equalisation, is the first step in the identification process. By applying the image classification region of the affected part, a Multiclass SVM can positively identify the sickness kind from the original leaf picture among 500 photographs with a 96% accuracy rate. Additionally, it shows the proportion of apple leaves that are diseased and infected.

S. Arya et al. [11] Employed the CNN and AlexNet architectures to examine the efficacy and accuracy of these architectures in identifying Mango and potato leaf sickness. For this study, a dataset with 4004 photos was employed. The findings demonstrate that AlexNet's accuracy outperforms CNN's architecture.

R. Indumathi et al. [12] Give a method that identifies both the disease that infected the leaf and the affected area of the leaf. Image processing is used to accomplish this; systems exist that foretell leaf illnesses. K-Medoid clustering and the Random Forest algorithm are used in the technique to increase the accuracy of leaf disease diagnosis. Before using the clustering technique, pre-processing is used to identify the leaf's impacted region.

U. Barman et al. [13] Provide a technique the SBCNN (Self-Built Convolutional Neural Network) for potato disease detection is also presented in this study. The datasets of potato leaves with and without enhancement are each subjected to a distinct application of the SBCNN. Testing and training of the potato leaf pictures are done using the algorithm. The greatest SBCNN training accuracy is 99.71%, while the best SBCNN validation accuracy in the non-enhanced and augmented datasets is 96.48%.

S. Kumar et al. [14] They have employed a variety of photos in the proposed effort to find leaf diseases. To separate different features, they used segmentation techniques like k-means clustering. Both the SVM (Support Vector Machine) classifier and the GLCM (Grey Level Co-occurrence Matrix) classifier are used to categories various illnesses. Using this method, they can precisely distinguish the various diseases in leaves. Images of different leaves that have been afflicted by ailments like Alternaria alternata, Cercospora leaf spot, bacterial blight, and anthracnose are included in the collection. The outcomes accurately display the area and the proportion of those impacted.

S. Ashok et al. [15] Proposed approach has a 98% accuracy rate. The suggested method maps input image pixel intensities and compares them with the trained dataset image using a CNN strategy for extracting features in hierarchies. By lowering the error relative to the training set, all changeable sections of the leaves' parameters are enhanced. The comparison image is identified using an image classifier technique that can also use artificial neural networks, fuzzy logic, and hybrid algorithms to distinguish between disease-affected and healthy leaves.

P. Sharma et al. [16] Describe a technique that uses a CNN (Convolutional Neural Network) for classification to present a very accurate artificial intelligence solution for identifying and categorizing various plant leaf diseases. The dataset utilized in the presented model has more than 20,000 photos divided into 19 different classes. The accuracy of the following model can be increased by fine-tuning the hyper parameters and utilizing an even larger dataset with additional disease categories. The model can also contain the treatments for the ailment that is classified. The model can then be made available on both the Android and iOS platforms to reach out to the farmers who can actually use the suggested system.

III. PROPOSED METHODOLOGY

Our proposed model main objective is to create a model for potato leaf disease detection that will help with disease identification. Data acquired from Kaggle is used in this job.

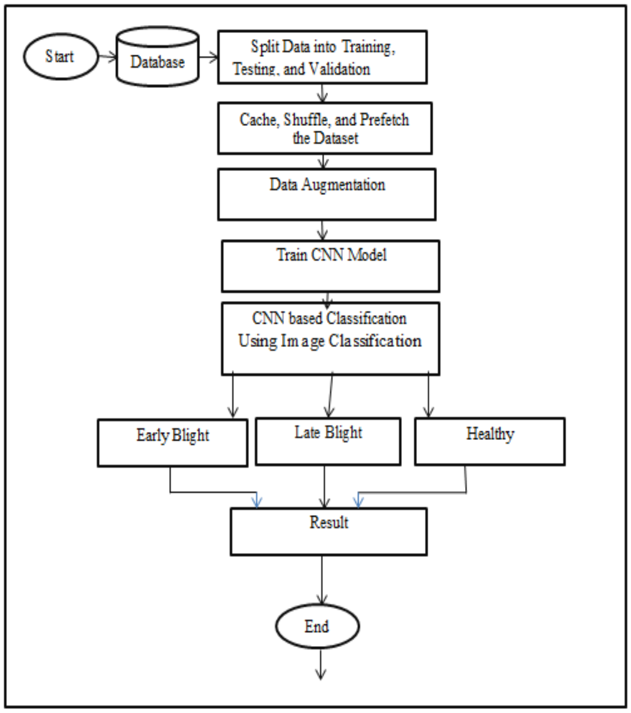

Figure.1 Block Diagram of the Potato Leaf Disease Detection System

Figure.1 represents the workflow of the potato leaf disease detection system. The dataset for the image consists of 2152 images in total, divided into 3 categories. The dataset comprises both healthy and diseased leaves and covers diseases including potato early blight and potato late blight. Each picture is in.jpg format.

A. Dataset

1. Dataset Example: Images of various potato leaf varieties from our dataset.

Figure.2 Image of Potato Leaf Infected by Early Blight Disease

Figure.3 Image of Potato Leaf Infected by Late Blight Disease

Figure.4 Image of Potato Healthy Leaf

2. Prepare the Training Data: The dataset size for experimental setup is 2152 number of images. The images of potato leaves will split into training, test and validation sets. The length of training dataset is 80%, length of dataset is 10% and the length of validation dataset is 10%. Table.1 shows number of images in potato leaf dataset.

Table.1 Number of Images in Potato Leaf Dataset

Name of the Plant	Disease Categories	Number of Images
Potato	Early Blight	1000
	Late Blight	1000
	Healthy	152

B. Cache, Shuffle, and Prefetch the Dataset:
After splitting data into training, testing and validation datasets. The next step is cache, shuffle and prefetch the dataset.

1. Cache:
After being loaded from disc during the first epoch, images are kept in memory. By doing this, you can be sure that the dataset won't slow down the training of your model. If your dataset is too big to store in memory, you may use this technique to quickly establish an on-disk cache.

2. Shuffle:
shuffle(buffer_size, seed=None, reshuffle_each_iteration=None, name=None
)
The components of the dataset are randomly combined. This dataset inserts elements in steps of buffer_size into the buffer, selects a random element from the buffer, and then loads new elements to replace the randomly chosen ones. For efficient shuffling, the buffer size must be more than or equal to the full dataset size.
For instance, shuffle will choose a random image from the top 1,000 photos in the buffer if buffer_size is set to 1,000 and our dataset has 2152 images. By swapping out an image that is selected with the image that comes after it (i.e., image 1,001), the buffer is preserved at 1,000 images.
3. Prefetch: Near the conclusion of the majority of dataset input pipelines, the prefetch call should be made. By doing so, it is possible to prepare next elements while the current element is being processed. As a result, there are frequently improvements in performance and latency at the price of higher memory usage to hold prefetched data. Prefetch functions similarly to other Dataset methods in that it manipulates the dataset's elements. It is unable to distinguish between batches and instances.

Cache, Shuffle, and Prefetch the Dataset

```
train_ds = train_ds.cache().shuffle(1000).prefetch(buffer_size=tf.data.AUTOTUNE)
val_ds = val_ds.cache().shuffle(1000).prefetch(buffer_size=tf.data.AUTOTUNE)
test_ds = test_ds.cache().shuffle(1000).prefetch(buffer_size=tf.data.AUTOTUNE)
```

Figure.5 Code of Cache, Shuffle and Prefetch Dataset in Our Proposed Model

C. Data Augmentation: When we have less data, we need to augment it. This increases the accuracy of our model by augmenting the data. It is necessary for building neural networks that can work in practical situations. By using data augmentation, we may improve our model's capacity to generalise and provide more precise predictions on data that wasn't used in its training.
The preprocessing module in TensorFlow and the Sequential class make it simple to include data augmentation into a tf.data pipeline. This technique is occasionally referred to as "layers data augmentation" since the Sequential class used for data augmentation is the same class used to build sequential neural networks (LeNet, VGGNet, AlexNet).
In our model, we are building a series of data augmentation activities such as:
1. Random horizontal and vertical flips
2. Random rotation

D. Train CNN Model: One of the most widely used methods in deep learning; convolutional neural networks (CNN), robustly trains many layers. It has been found to be quite effective and is also the one that computer vision programmers use the most. A computational framework that converts unstructured image inputs into the required output groups for categorization can be built using CNN [6]. In our research work, the CNN model is implemented on the Potato dataset. With a batch size of 32, the model is compiled for 50 epochs. These are the steps in this algorithm:

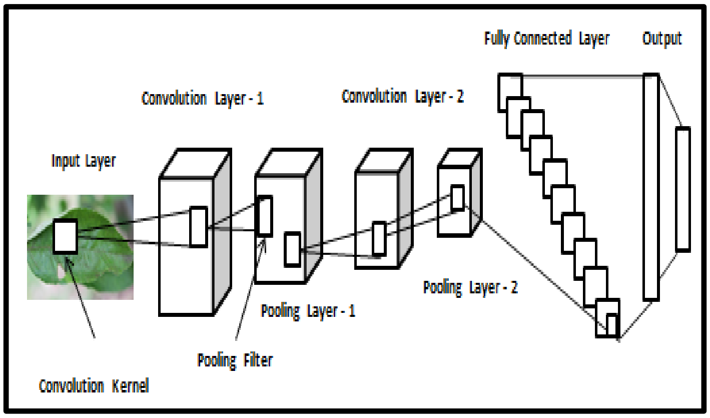

Figure.6 CNN Model Architecture

1. Image pre-processing

The dataset's images have been shrunk to 256 x 256 sizes. As a result, training time can be reduced and a model that is computationally realistic can be created.The training process can be accelerated by optimising the input or goal variables. while protecting the image's data integrity from loss.

2. Input layer

The source pictures and their pixel numbers are located in the input layer.

3. Convolution layer

Different feature maps are produced by a CNN by converting both the best feature maps and the complete object using various kernels in the convolution layers. Six convolution layers were utilised in our investigation. Each layer is 2D. The model's first Convolution layer has a layer size of 32. The first layer's kernel size is 3X3 and the activation function Relu. The model's second Convolution layer has a layer size of 64. The second layer's kernel size is 3X3, and the activation function Relu. The model's third Convolution layer has a layer size of 64. The third layer's kernel size is 3X3, and the activation function Relu. The model's fourth Convolution layer has a layer size of 64. The fourth layer's kernel size is 3X3, and the activation function Relu. The model's fifth Convolution layer has a layer size of 64. The fifth layer's kernel size is 3X3, and the activation function Relu. The model's sixth Convolution layer has a layer size of 64. The sixth layer's kernel size is 3X3, and the activation function Relu.

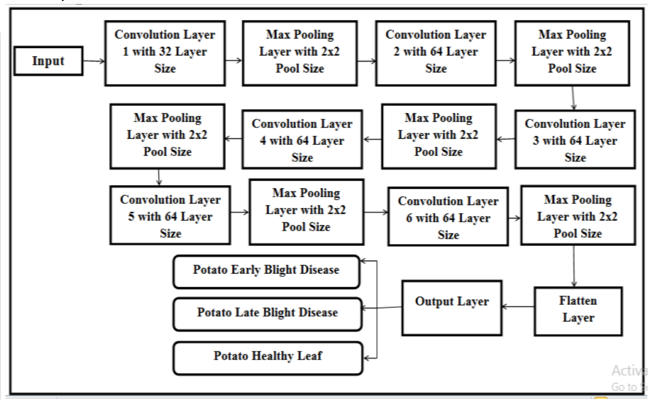

Figure.7 CNN Model Architecture for Potato Leaf Disease

4. Pooling layer

A pooling layer, which may be used to compress feature maps and network parameters, is typically added after convolutional layers in machine learning algorithms.

Because they factor in nearby pixels in their calculations, pooling layers are similarly invariant in interpretation to convolutional layers. The layer's pool size is 2x2.

5. Non-linear layer

The convolution neural network transforms the input in a non-linear way with the goal of classifying the characteristics in each hidden layer. We utilise Rectified linear units in the CNN structure (ReLU). Non-linear transformations frequently employ rectified linear units.

6. Fully connected layer

The fully connected node, the last layer of the convolution neural network, is applied to the data after several cycles of the other layers.

The neurons in the fully interconnected network are physically connected to the neurons in the two neighboring layers.

7. Softmax layer

The output layer is augmented with three hidden neurons using a SoftMax activation function to forecast the three kinds of potato illnesses.

8. Training and testing

A network is trained by assigning weights to fully connected layers and kernels to convolutional layers that minimize the differences between output predictions and provided ground truth classifications in a training dataset.

We used 80% of the data in our study for training, 10% for testing, and 10% for validation.

E. Compiling the Model: putting "adam" to work as a useful optimizer. As the training session goes on, this is a useful tool for improving the learning rate. We use "categorical cross-entropy" to train the system, and we provide the "accuracy" metric to indicate the accuracy score on the validation set.

IV. EXPERIMENT RESULT

We use 2152 pictures of potato leaves to our algorithm. They were all 256 × 256 pixels in size.2000 sick leaves and 152 healthy leaves make our dataset. We used 80% of the data in our study for training, 10% for testing, and 10% for validation [8].

The CNN result for potato leaves is explained here.

1. Training Process

We mix a CNN with softmax activation in the output layer. The underlying layers for data augmentation, normalisation, and scalability are also covered. Images are typically categorized using CNN. For the duration of the research, most there were 64 and 50 epochs in each batch. Each epoch's beginning value must match the picture being learnt. To determine the values for loss and accuracy, the epoch results will be recorded. While the loss value—which must be close to or equal to zero—shows how poorly the model has done, the accuracy number reveals how accurately the system categorizes items [5].

The graph (Figure.9) indicates that the CNN's training and validation accuracy for potato leaves increases with epochs.

```
acc = history.history['accuracy']
val_acc = history.history['val_accuracy']

loss = history.history['loss']
val_loss = history.history['val_loss']
```

```
plt.figure(figsize=(8, 8))
plt.subplot(1, 2, 1)
plt.plot(range(EPOCHS), acc, label='Training Accuracy')
plt.plot(range(EPOCHS), val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation Accuracy')

plt.subplot(1, 2, 2)
plt.plot(range(EPOCHS), loss, label='Training Loss')
plt.plot(range(EPOCHS), val_loss, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')
plt.show()
```

Figure.8 Code of Graphical Plot of CNN for Potato Leaves in Our Proposed Model

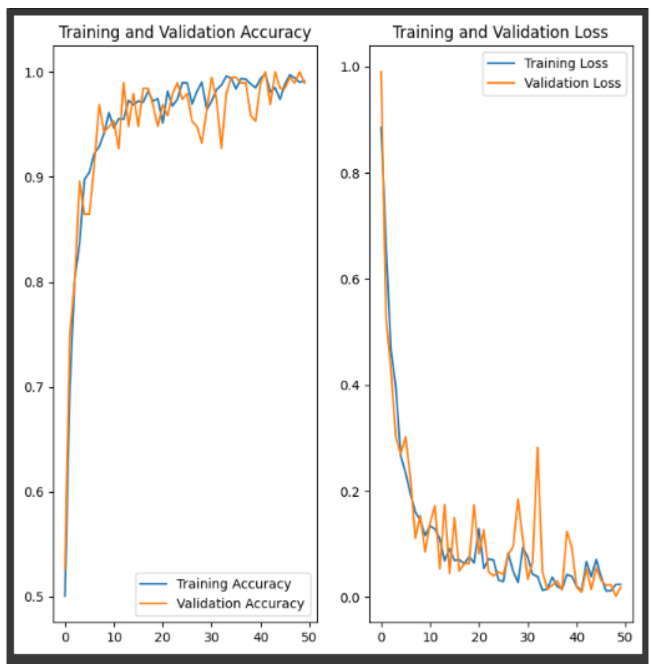

Figure.9 Graphical Plot of CNN for Potato Leaves

2. Testing Process

After training on the acquired datasets, we extract additional data from outside the training datasets to test with the CNN model. The whole process is implemented in the Google colab.

Following is our predication:

```
import numpy as np
for images_batch, labels_batch in test_ds.take(1):

    first_image = images_batch[0].numpy().astype('uint8')
    first_label = labels_batch[0].numpy()

    print("first image to predict")
    plt.imshow(first_image)
    print("actual label:",class_names[first_label])

    batch_prediction = model.predict(images_batch)
    print("predicted label:",class_names[np.argmax(batch_prediction[0])])
```

```
first image to predict
actual label: Potato___Early_blight
predicted label: Potato___Early_blight
```

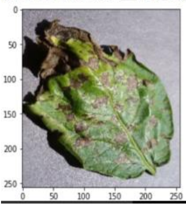

Figure.10 Code for Predication of Potato Leaf Disease in Our Proposed Model

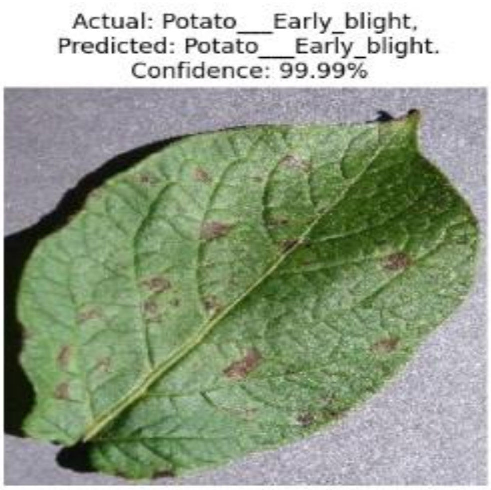

Figure.11 Potato Leaf affected by Early Blight Disease

Figure.11 depicts a leaf affected by Early Blight. The dots are frequently surrounded by a golden or striking green-yellow circle. Larger spots may eventually cause the yellowing and death of the entire leaf.

Figure.12 Potato Leaf affected by Late Blight Disease

Figure.12 depicts the leaves being affected by late blight. Small, irregularly shaped green particles that range in hue from dull to dark green make up the earliest leaf flecks. The specks quickly grow into large brown to purple-black areas in cold, dry conditions. The entire leaf may get infected by the illness, which has the capacity to infect stem tips and destroy the entire plant.

Figure.13 Healthy Potato Leaf

Figure.13 depicts a healthy potato leaf. As a result, our provided model accurately distinguishes between sick and healthy photos.

V. CONCLUSION

The proposed approach gained a 98% accuracy level. As potato disease identification, we used a CNN (convolution neural network) model based on the classification technique known as the legitimate sequential model. The research indicates that data augmentation is an effective strategy for improving the performance of the CNN model. The provided model correctly distinguishes between damaged and healthy plants. We just listed two types of illnesses, yet this is insufficient for farmers or those involved in the agricultural industry. In order to get the best results, we chose to raise the total number of diseases to make it more probable for them as well as the amount of picture datasets.

REFERENCES

[1] Xin Li, Laxmisha Rai (2020), “Apple leaf disease identification and classification using Resnet models,” 2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT), pp. 738–742.

[2] V. V. Srinidhi, A. Sahay and K. Deeba, "Plant Pathology Disease Detection in Apple Leaves Using Deep Convolutional Neural Networks : Apple Leaves Disease Detection using EfficientNet and DenseNet," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1119-1127

[3] K. Ahmed, T. R. Shahidi, S. M. Irfanul Alam and S. Momen, "Rice Leaf Disease Detection Using Machine Learning Techniques," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 2019, pp. 1-5

[4] S. V. Militante, B. D. Gerardo and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2019, pp. 579-582

[5] R. A. Sholihati, I. A. Sulistijono, A. Risnumawan and E. Kusumawati, "Potato Leaf Disease Classification Using Deep Learning Approach," 2020 International Electronics Symposium (IES), 2020, pp. 392-397

[6] J. M. Al-Tuwaijari, M. A. Jasim and M. A. - B. Raheem, "Deep Learning Techniques Toward Advancement of Plant Leaf Diseases Detection," 2020 2nd Al-Noor International Conference for Science and Technology (NICST), 2020, pp. 7-12

[7] M. K. R. Asif, M. A. Rahman and M. H. Hena, "CNN based Disease Detection Approach on Potato Leaves," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 428-432

[8] S. Chakraborty, S. Paul and M. Rahat-uz-Zaman, "Prediction of Apple Leaf Diseases Using Multiclass Support Vector Machine," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 147-151

[9] M. V. Applalanaidu and G. Kumaravelan, "A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 716-724

[10] L. S. Puspha Annabel, T. Annapoorani and P. Deepalakshmi, "Machine Learning for Plant Leaf Disease Detection and Classification – A Review," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0538-0542

[11] S. Arya and R. Singh, "A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf," 2019 International Conference on Issues and

Challenges in Intelligent Computing Techniques (ICICT), 2019, pp. 1-6

[12] R. Indumathi., N. Saagari., V. Thejuswini. and R. Swarnareka., "Leaf Disease Detection and Fertilizer Suggestion," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019, pp. 1-7

[13] U. Barman, D. Sahu, G. G. Barman and J. Das, "Comparative Assessment of Deep Learning to Detect the Leaf Diseases of Potato based on Data Augmentation," 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 682-687

[14] S. Kumar, K. Prasad, A. Srilekha, T. Suman, B. P. Rao and J. N. Vamshi Krishna, "Leaf Disease Detection and Classification based on Machine Learning," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 361-365

[15] S. Ashok, G. Kishore, V. Rajesh, S. Suchitras, S. G. G. Sophia and B. Pavithra, "Tomato Leaf Disease Detection Using Deep Learning Techniques," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 979-983

[16] P. Sharma, P. Hans and S. C. Gupta, "Classification Of Plant Leaf Diseases Using Machine Learning And Image Preprocessing Techniques," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 480-484