POTATO LEAF DISEASE CLASSIFICATION USING CNN

Dr. M. Janga Reddy, Mr. P. Pavan Kumar, Ms. B. Sunitha Devi, Mr. Katakam Srinivasa Rao, Dr. D. Baswaraj, Mr. M. Ravi, CMR Institute of Technology Hyderabad

Abstract

Annually, a significant amount of the potato crop is lost to a number of illnesses, including blight. Blight infects the plant's leaves, causing them to rot and wither, causing the plant to die by interfering with photosynthesis. If the disease's effects are widespread, this could cause the agriculturalists to suffer severe losses. Because of advances in Deep Learning and Artificial Intelligence in irrigation, rot can now be spotted in potato leaves, which, if detected early, can help to minimise devastation. For the purpose of identifying diseases in various plants, numerous models have been developed in the past. A CNN (Convolutional Neural Network) model was developed and evaluated in this article to detect and retrieve significant data from a dataset and use it to evaluate whether a potato leaf is healthy, has late blight, or has early blight. On the test dataset, this technique had an accuracy of 97.92%.

Keywords: deep learning; CNN; blight; potato.

1. INTRODUCTION

In India, agriculture is the main industry. Approximately 60% of the workforce in India is employed in the agriculture industry, which is one of the important industries in the nation. Over 30% of all crops grown in India are potatoes, one of the most well-liked and versatile plants [1]. This is a fourth-most-grown eatery crop worldwide and the second- most-grown crop in India.. India harvested 53.69 million metric tonnes of potatos in 2021. This represents an increase of more than 5 millions metric tonnes over the harvest in 2020. Farmers who also grew potatos suffer notable money losses annually as a result of the various diseases that can affect a potato plant. Therefore, an accurate disease detection system is essential for effective crop management that prevents the spread of illnesses.

Bacteria ring decay, early fungus,and late fungus are some prevalent diseases. Clavibacter michiganensis, a type of bacterium, is responsible for bacterial ring rot, which is characterised by curled leafs, fading of the plant, and tuber rotting. Bacterial wilt is another such illness that causes the entire plant to wilt and die once the leaves become yellow at their bases. When stems are sliced, a brown ring will be evident. Microorganisms, including things like bacteria, fungus, and viruses, are the primary culprits of the majority of these disorders.

Alternaria solani is a type of fungus that causes early blight, and Phytophthora infestans, an oomycete, causes late blight. If a agriculturalist can identify these illness fast and take the necessary measures to treat them, they can avoid significant crop losses and financial damages [2]. It's critical to correctly identify the type of disease affecting that potato plant because treatments for early blight differ from those for late blight.

Potato Early Blight

Previously, conventional image procesing technique such as K-means clustering were commonly recommended and used for diagnosing these leaf diseases. CNN (Convolutional Neural Network) was used in our proposed journal paper to evaluate photos and procesed them to determine whether they had fungus or if the plant was healthy [3]. Other authors have proposed a variety of deep learning methods.

Potato Late Blight

2. LITERATURE REVIEW

Throughout the term of this project, multiple deep learning algorithms to identify leaf illnesses in potato plants from various researchers were discovered through their separate research publications. A synopsis of those research publications is provided in this section. The classification of plant diseases has been the subject of research using a variety of techniques. However, it is still regarded as missing [5] and has evolved into a study topic that is currently being explored due to how widely this field's subject matter differs. Agriculture-related articles totaling 37 were released between 2015 and 2017. More specifically, 7 papers in 2016,
15 papers in 2017, and 15 papers in 2016 were published. This fact demonstrates how recent and cutting-edge this technology is in the field of agriculture [6].

It is possible to create a deep learning-based method for learning an usable features representation. Several occipital cortex tasks, including text detection, victim detection, target tracking, and object detection, have shown great result using deep learning [7]. Increasing accuracy requires a deeper network. It's noteworthy to observe that the deeper layers of the deep netvork are more robust to changes in location, colour, scale, and flexible object because they can record more semantic information or conceptualization. They may therefore be suitable for properly diagnosing leaf diseases.

To recognise and classify Septoriea Leaf Spot and Yelloow Leaf Currl, two tomatoa leaf illness that are frequently founded in tomatoa harvests, Prachi. [3] used the Leneet architectural technique and Convolutional Neural Network model topology. Through PlantVillage, one of the freely available image databases, researchers had access to the dataset. Predictions made using the method recommended for this study are 90% to 93% accurate.

Convolutional neural network (CNN) technology is used to evaluate the Sedan Classification refers [17] model and identify 10 different diseases in five distinct plant species. The total outcome of this analysis was 96.3%. Using the same techniques but a different design and the U-net architectural, Caroline Fuji al et. [18] built a viral health surveillance system for cucumber plants, and they were successful in achieving an average efficacy of 80.3%.

Raj M & colleagues. [19] collected 5000 photos of tomato leaves from the Plant Village accessible image database. The dataset is leveraged, together with a number of characteristics and the Classifiers SVM (Support Vector Machine) method, to describe the three potential illnesses that may affect maise leaves as objects. In this study, the efficiency of the technique was compared to that of the statistical approach and device grey level founder. The gathered data demonstrates the total accuracy of the SVM (Support Vector Machine) Multi technique. Segmented with K-mean grouping is a technique used by Panjali M. Padol and Prof. Kamala K. Jadav [20] to determine the illness region and then determine the colors and textures of the image. They obtain a classification using the SVM (Svm Classifier) Encoder technique. performance of 80.89%.

To characterize various plants illness kinds, Jeol Joseph et al. [21] used the Backpropagation Neural Network approach using the GLCM image retrieval stage. Berry Sorts Of diseases, a type of staph infection that affects peppers, and Accelerated Illness, a disease brought on by a deficiency in minerals like nitrogen, magnesium, and potassium, are two examples of pepper plant pathogens that can be observed. This classification system was used by Eftekhar Hossain et al. [22] to classify illnesses including Alterrnaria Alternrata, Anthrracnose, Bacterieal Rot, Leaf Spot, and Plant Leafve Cancer. In this investigation, the KNN disease diagnosis approach has a 98.88% accuracy rate.

Artificial neural networks and fuzzy logic techniques were used in a study by Aakanksha Rastogi et al. [23] to identify plant illnesses based on leaf characteristics. The goal of this study is to recognise and categorise leaf diseases in maple and hydrangea plants. There are two classifications for the disease-affected plants' leaves: scorched leaves and leaf spots. Leaf spots are areas of the leaf where the disease is concentrated, whereas burnt leaves are areas of the leaf where the disease has spread widely.

A pre-trained Masked RCNN technique based on the Microsofts Comon Thing in Contexxt (MSCOCO) datasample was used in a study by Srikanth Ramanujan. This model is based on domain adaptation or datasets with outdated information. This model provides a 92 % optimistic accuracy rate. [24] Priyanka Shetty was successful in extracting the suitable and beneficial features from the dataset using the ODMT, the VGG architecture, and the CNN model outlined in the study. This makes use of the open-source Python programme Orange Data Mining. The technique's overall accuracy was 95.8%.

3. PROPOSED METHODLOGY

A. Tools and Libraries

1) Tools
    i) Anaconda Navigator
    ii) Jupyter Notebook
2) Libraries
    i) Tensorflow-vers-2.3
    ii) Keras-vers-2.4
    iii) Matplotlib-vers-3.4
    iv) Numpy-vers-1.19

B. Dataset Description

This model is trained, tested, and validated using a Kaggle data set. Three classes, Early blight, Late blight, and healthy, were established, with 1000, 1000, and 152 photos in each class, respectively.

Category	Total Samples	Training Images	Validation Images	Test Image
Early blight	1000	800	100	100
Late blight	1000	800	100	100
Healthy	152	122	15	15
Total	2152	1722	215	215

C. Data Preprocessing

The technique of removing extraneous data from images is known as data pre-processing. We do not use photographs with huge quantities of unnecessary data. To normalise the input photographs in the dataset, samples from numerous sources with varying sizes will be transformed to 256x256 px.

D. Data Augmentation

Data Augmentation is a technique for quickly and effectively increasing the size of a training instance by creating new data from previously gathered information. Consider using Data Augmentation if you want to prevent data dimensionality, if the existing data appears to be too small to train, or if you still want to squeeze more effectiveness out of our prototype.

To really be straightforward, data augmentation is used for purposes other than avoiding fitting problem. A large dataset is essential for both ML and Dl models to perform well. Even so, by augmenting the information we already have, we can improve the effectiveness of the algorithm. This implies that Data Augmentation can significantly boost the algorithm’s effectiveness.

E. Training the Algorithm

To maximize model performance, a sequential technique with a scaling layer and normalisation layer is built before feeding the sample images to the CNN. All photos are resized by the resizing layer to 256x256 pixels in size. The normalisation layer adjusts the pixel value of the image (keeping them in range 0 and 1 by dividing by 256).

The CNN layer is compensated for by convolutional layers, max pooling 2D layers, and dense layers. The actual filter size is 64 and the number of convolutional layers is 64. (3,3). The activation function receives the end product of convolution neurons. This activation function broadens the network's abilities by allowing it to learn large amounts of detailed data and express non-linear complicated arbitrary functional mappings between inputs and outputs. As a result, we were able to construct nonlinear mappings from inputs to outputs using nonlinear activation. In our case, the activation function is provided by ReLU layers. These layers will eventually be compressed to form an array of convolutional nodes. neurons.

The max pooling layer is needed to preserve the attributes, and the Dense layer implies that the image size will be lowered by the preceding layer's weights. The use of as the input for all neurons in the layer below is referred to as "max-pooling." This layering method for extracting a matrix's highest value. We currently have 64 neurons. The number of convolutional layers to form a stack of, as well as the max-pooling classes with softmax activation function, are contained in the final layer.

4. RESULT

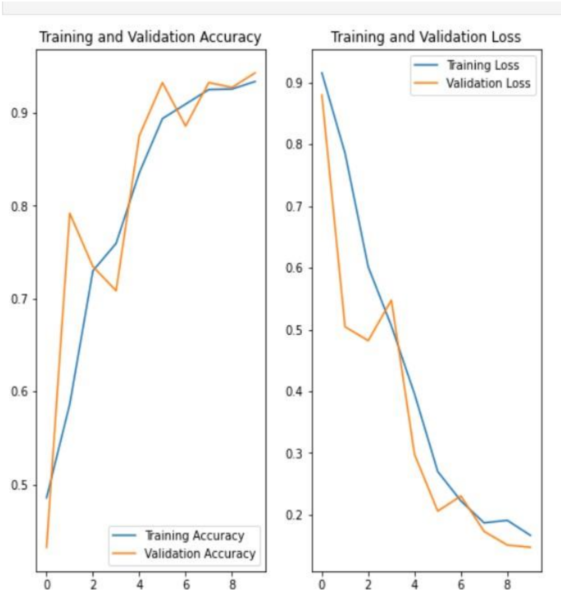
2152 photos of potato leaf’s were gathered from various websites and enhanced with data to produce a final dataset of 2152 images for the model presented here. 80 percent of the data set was used for training (1900 photos), ten percent for validation (100 photos), and ten percent for testing (52 pics). CNN (Convolutional Neural Network) was used to extract the data needed from those photos to classified them as early, late, or health blight.

The final accuracy of training is 98.94%, while the accuracy for validation was 100%, with minimal losses of 0.17% and 0.31%, respectively, as shown in fig.5. Figure 4 depicts epochs as well as the development of the model training. The test examples used to validate the model after training are shown in Figure 8. The system's accuracy increased from previous studies by 2.04% to 99.84% after it was successfully implemented. Our method can help farmers increase crop yields while decreasing crop blight damage.

```
54/54 [==========================] - 92s 2s/step - loss: 0.9155 - accuracy: 0.4855 - val_loss: 0.8796 - val_accuracy: 0.4323
Epoch 2/10
54/54 [==========================] - 77s 1s/step - loss: 0.7855 - accuracy: 0.5862 - val_loss: 0.5044 - val_accuracy: 0.7917
Epoch 3/10
54/54 [==========================] - 75s 1s/step - loss: 0.6015 - accuracy: 0.7297 - val_loss: 0.4819 - val_accuracy: 0.7344
Epoch 4/10
54/54 [==========================] - 76s 1s/step - loss: 0.5057 - accuracy: 0.7593 - val_loss: 0.5476 - val_accuracy: 0.7083
Epoch 5/10
54/54 [==========================] - 76s 1s/step - loss: 0.3955 - accuracy: 0.8351 - val_loss: 0.2975 - val_accuracy: 0.8750
Epoch 6/10
54/54 [==========================] - 78s 1s/step - loss: 0.2700 - accuracy: 0.8935 - val_loss: 0.2060 - val_accuracy: 0.9323
Epoch 7/10
54/54 [==========================] - 76s 1s/step - loss: 0.2225 - accuracy: 0.9091 - val_loss: 0.2304 - val_accuracy: 0.8854
Epoch 8/10
54/54 [==========================] - 76s 1s/step - loss: 0.1869 - accuracy: 0.9248 - val_loss: 0.1732 - val_accuracy: 0.9323
Epoch 9/10
54/54 [==========================] - 76s 1s/step - loss: 0.1910 - accuracy: 0.9253 - val_loss: 0.1511 - val_accuracy: 0.9271
Epoch 10/10
54/54 [==========================] - 76s 1s/step - loss: 0.1667 - accuracy: 0.9334 - val_loss: 0.1473 - val_accuracy: 0.9427
```

Training the Model

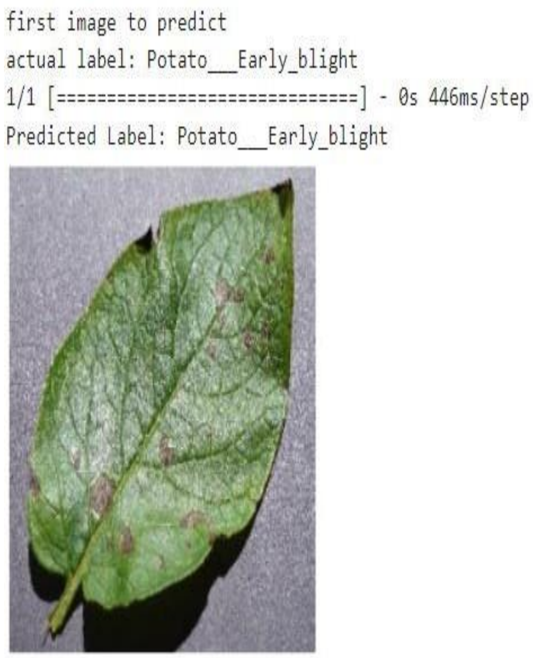
CNN was used in this study to improve the model's accuracy. Epoch-epoch, the neural network of training data is trained forward and backward once. Epcoh all of the training data at once. One pass refers to both forward and backward movement. Epoch includes multiple batches, each of which contains a portion of the dataset used to train the neural network. An epoch is one process of the total training set in the framework of artificial neural networks. Training a neural network normally takes several epochs. In other statements, if we serve a neural network training data in various of patterns over multiple epochs, we can expect gretaer sweeping generalization when confronted with new "hidden" input (testing data). In real-world scenarios, the model's results would be less accurate if the number of epochs was reduced, rendering it useless for identifying blight in potato leaves.



Accuracy and Loss for Training and model

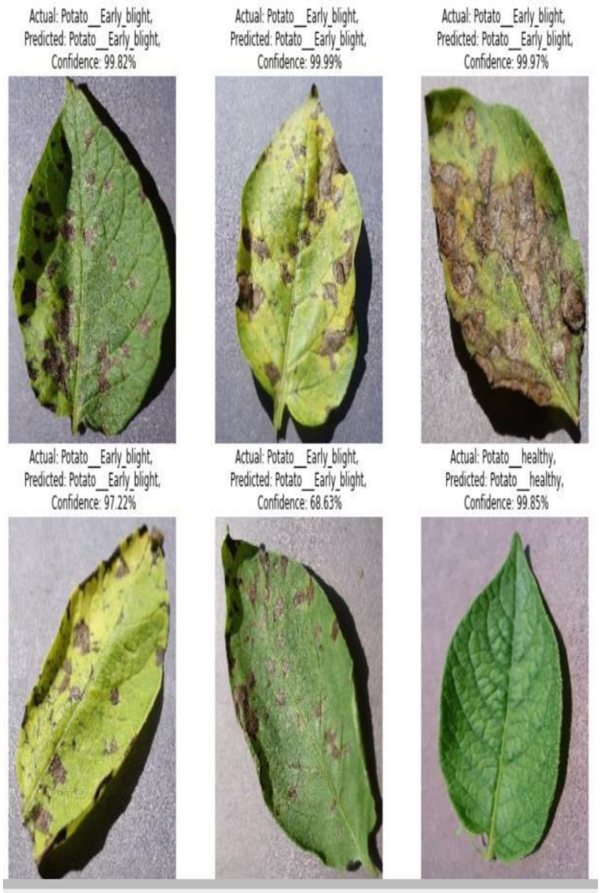
As an outcome, we chosen 5 epochs in the training and testing situations that yielded the maximum accuracy with the least amount of data loss (0.15% and 0.21%, respectively). The

CNN's hidden layer design was built using convolutional layers, max pooling 2D layers, and thick layers. The actual filter size of the convolutional layers (a total of 64 layers) is (3,3). These layers are then passed on to the activation layer, which helps the network learn more complex facts.



Testing

REFERENCES



5. FUTURE SCOPE

Prediction

[1] K. . A. Beals, "Potatoes, Nutrition and Health," American Journal of Potato Research, no. 96, pp. 102-110, 2019.
[2] R. K. Arora and S. Sharma, "Pre and Post Harvest Diseases of Potato and Their Management," in Future Challenges in Crop Protection Against Fungal Pathogens (Eds. Aakash Goyal and C.Manoharachary), New York, Springer, 2014, pp. 149-183.
[3] P. TM, P. Alla, K. S. Ashirta, N. B. Chittaragi and S. G. Koolagudi, "Tomato Leaf Disease Detection using Convolutional Neural Network," International Conference on Contemporary Computing (IC3), 2018. [4] "An Updated Survey of Efficient Hardware Architectures for Accelerating Deep Convolutional Neural Networks," Future Internet, vol. 12, no. 113, pp. 1-22, 2020.
[5] S. Sladojevic, M. Arsenovic, A. Andras, D. Culibrk and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," Computational Intelligence and Neuroscience, p. 11, 2016.
[6] "Enhanced Field-Based Detection of Potato Blight in Complex Backgrounds Using Deep Learning," Joe Johnson, Geetanjali Sharma, Srikant Srinivasan, Shyam Singh, Sanjay Jadhav, Jaydev Sarma, and Vinay K D, 2021.

This approach would be incorporated and made easily accessible to everyone through a web-based app. This would be created with a comparatively modern framwork, like React or Angular; this developments would be utilised to create the front end of the programme (UI). The user interface (UI) will present good explanations to the exact stage of blight identified, supporting the user in protecting his crops.

Flask, FastAPI, Django REST framework, and other technologies could be used to link the frontend and the machine learning model in the backend. Additionally, treatments for the blight would be suggested based on the state of the plant at the time.

6. ACKNOWLEDMENT

This research has been guided by Ms. T. Gayathri, Assistant Professor, Department of Information Technology, Vignan’s Foundation for Sciences, Technology and Research.

[7] "Potato Leaf Disease Detection Using Inception V3", Volume: 07 Issue: 10, November 2021, by Gunjan Saxena, Amar Singh, Prathap Chowdhary, and Ruhali Shetty
[8] "Potato Plant Disease Detection Using Convolution Neural Network", Parmesh R, Sharath Sharma M, Santhosh Reddy D, Madhu Grover CH, 2020.
[9] Plant Disease Classification Using Deep Learning Google Net Mode, Sathwinder Kaif, Garisma Joshe, and Renuka Varma, 2019.
[10] K. Yasodha and A. Krishna Chaitanya, "Image based Plant Disease Detection using Convolution Neural Networks Algorithm", 2021.
[11] List of nations by the amount of potatoes they produce: https://en.wikipedia.or/wiki/List of nations by the amount of potatoes