Crop Disease Detection Using Image Analysis

Dr. M. Mohamed Iqbal
Assistant Professor Senior Grade 1
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
iqbalmecse@gmail.com

R. Jyothsna
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh

S.P. Deekshita
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh

Y. Renuka Gayathri
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh

Abstract **Crop diseases pose a severe danger to global food security, therefore for successful crop production, it is crucial to find the disease early on. Early disease diagnosis and crop health information can make it easier to manage illnesses through effective management techniques. Crop productivity will increase as a result of this strategy. Farmers typically examine the plants to spot infections. As an alternative, we can utilize automatic methods for classifying plant illnesses, which aid in acting after seeing leaf disease symptoms. A crop disease prediction system using a Convolutional neural network model is offered in this research. The methodology is to train a deep convolutional neural network to recognize 6 diseases using a plant village dataset of 6000 photos of damaged and healthy plant leaves taken under controlled conditions. The trained model's 95% accuracy on a held-out test set demonstrates the viability of this method.**

Keywords **Crop diseases, Disease Detection, Deep Learning, Tensorflow, Convolutional Neural Network**

I. INTRODUCTION

The primary source of food, raw materials, and fuel that supports a country's economic growth is Agriculture. As the world's population is expanding quickly, agriculture is finding it difficult to meet all its needs. Numerous factors, such as climate change, the reduction of pollinators, crop diseases, a lack of irrigation, etc., continue to pose a danger to food security. The ability to control agricultural illnesses by spotting them as soon as they appear on crops is a benefit. The cultivation of crops that are contaminated by illnesses is one of the biggest difficulties that farmers in rural areas encounter. Crop diseases have the potential to disrupt and harm both the ecological and economic equilibrium. It takes a lot of time and effort to monitor the farm and identify the various illnesses that afflict plants as a result of the farms [1]. Plant diseases (like mosaic virus, late blight, tomato bacteria, early blight, and yellow curved) have recently become more prevalent and are negatively influencing plant growth as well as the quality and quantity of production. Infections with viruses, bacteria, or fungi, for example, can be produced by insects or by any other element of nature. To achieve the right treatments and a better, quicker yield, these must be properly identified.

This conference paper aims to present a comprehensive exploration of crop disease detection using image analysis techniques. The purpose of this paper is to present an application that, using leaf textural similarity, predicts the type of crop disease. The algorithm is trained using a dataset from a plant community that includes both healthy and damaged crop leaves. Early detection of agricultural diseases can be used to stop additional crop damage, which is beneficial for maintaining farming.

In the following sections, we will review the key components of crop disease detection using image analysis. The paper will cover data collection and preprocessing techniques, essential for obtaining high-quality images representative of different crop diseases and environmental conditions. Feature extraction methods will be discussed, as they play a crucial role in capturing unique disease-related patterns from the images.

II. LITERATURE SURVEY

Crop disease detection has seen significant advancements with the application of deep learning and machine learning techniques. Convolutional Neural Networks (CNNs) have shown great promise in automatically learning relevant features from images and achieving high accuracy in disease identification. The literature often discusses various image-based techniques for crop disease detection, including color-based methods, texture analysis, and shape-based approaches. In [2] it can be observed how the accuracy has increased when the architecture has changed from Alex Net 85.53% to Google Net 99.34%.

Hyperspectral and multispectral imaging are gaining popularity for crop disease detection due to their ability to capture a wide range of spectral information [3]. Images that have been enhanced have higher quality and clarity than the original image. The segmentation of the leaf picture is critical when extracting features from that image. In [4], in addition to image analysis, the model focuses on using other sensors, such as thermal cameras, to detect crop diseases based on temperature variations associated with infection. In [5] the use of hyperspectral and multispectral imaging, coupled with advancements in drone and UAV technology, has opened new avenues for disease detection. These remote sensing methods offer the potential for early disease detection before visible symptoms manifest, enabling proactive disease management and resource allocation.

III. PROPOSED METHODOLOGY

This part discusses the classification of the model into six classes using leaf images. A model created to detect disease in crops at an early stage so that certain preventive measures can be taken before the disease spreads completely. The proposed model comprises five major components: The sequential layer, the Convolutional layer, the Pooling layer, flatten layer, and the dense layer.

A. Dataset Description

The dataset used is an open-source dataset called the PlantVillage dataset collected from Kaggle [6] which consists of diseased leaf images and corresponding labels. This dataset consists of 20,639 images spread into 15 labels of size 256 x 256 pixels.

Across all the experiment, only the first six classes are considered of the whole dataset for better accuracy. Figure 1 shows one image each with a label from the PlantVillage dataset. The batch size is taken as 32 and the dataset is partitioned into train, test, and validation sets. It must be noted that the model should perform better on unseen data and to avoid overfitting dataset is divided according to 80% training, 10% validation, and 10% testing respectively.

Table 1 shows a description of the entire dataset.

Table I – Brief Description of the Dataset

Disease Class	**Train**	**Test**	**Validation**
Potato_Early blight	800	100	100
Potato healthy	122	15	15
Potato_Late blight	800	100	100
Tomato heathy	1273	159	159
Tomato_virus	299	37	37
Tomato Bacterial_spot	1702	212	212

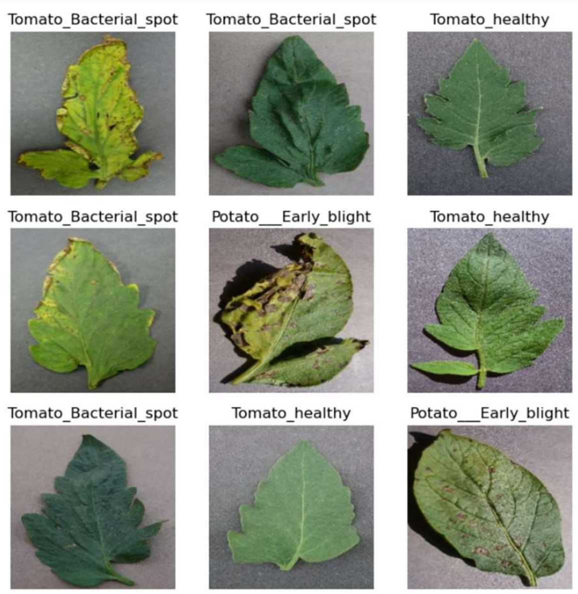

Fig. 1. Visualization of images from each class

B. Data Pre-processing

Before building a model, preprocessing is necessary. From the Tensorflow package, using Keras preprocessing layers, the dataset batches can be resized and rescaled. Normalization can help the training of our neural networks to rescale data such that it is between 0 and 1.

The formula for Standard Scaling:

$$z = (x - u) / s$$

This method normalizes each feature by removing the mean(u) of the training data and dividing by standard deviation(s). Data augmentation is also very useful for training neural networks. By applying this, more accurate predictions can be made on unseen data. Using Keras preprocessing layers for data augmentation can be done like layers.RandomFlip and layers.Random rotation over the same image repeatedly.

C. Building a model

To construct a model, a sequential CNN model with a convolution layer of size (5,5), a MaxPooling layer of size (2,2), nodes that have been flattened, and a fully connected layer with the dense layer's activation function being ReLu, which categorizes the six labeled objects, is used. Figure 2 shows an implementation of the CNN model.

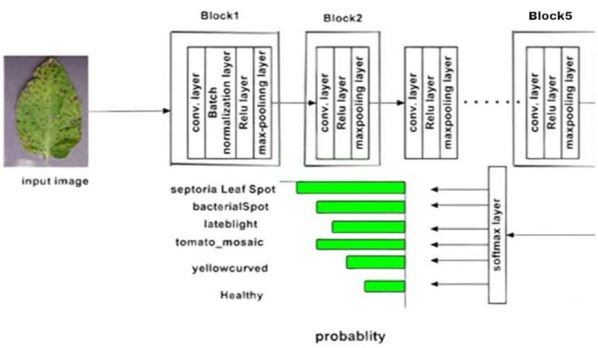

Fig. 2. Implementation of CNN model

D. Model Architecture

Figure 4 shows a detailed explanation of the model. The size of the input is (32,256,256,3) and the dataset is divided into 32 batches. The first layer is a Sequential Layer and creates no parameters since it is a sequential layer. And then a Conv2D layer produces an output of size (32, 256, 256, 32) and it has 896 parameters. The next layer max_pooling does not create any parameters. After pooling, flatten step is applied to flatten the connection of the model. Next, to flatten the layer, there are two dense layers. The dense layer is also known as the fully connected layer.

The suggested model has about 278,086 parameters.

After executing this model, it has been observed that the model is giving an accuracy of 95.0 on the training dataset.

```
model.summary()

Model: "sequential_6"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 sequential (Sequential)     (None, 256, 256, None)    0

 sequential_1 (Sequential)   (None, 256, 256, None)    0

 conv2d_20 (Conv2D)          (32, 254, 254, 32)        896

 max_pooling2d_20 (MaxPooli  (32, 127, 127, 32)        0
 ng2D)

 conv2d_21 (Conv2D)          (32, 125, 125, 64)        18496

 max_pooling2d_21 (MaxPooli  (32, 62, 62, 64)          0
 ng2D)

 conv2d_22 (Conv2D)          (32, 60, 60, 64)          36928

 max_pooling2d_22 (MaxPooli  (32, 30, 30, 64)          0
 ng2D)

 conv2d_23 (Conv2D)          (32, 28, 28, 64)          36928

 max_pooling2d_23 (MaxPooli  (32, 14, 14, 64)          0
 ng2D)

 conv2d_24 (Conv2D)          (32, 12, 12, 64)          36928

 max_pooling2d_24 (MaxPooli  (32, 6, 6, 64)            0
 ng2D)

 flatten_4 (Flatten)         (32, 2304)                0

 dense_8 (Dense)             (32, 64)                  147520

 dense_9 (Dense)             (32, 6)                   390

=================================================================
Total params: 278086 (1.06 MB)
Trainable params: 278086 (1.06 MB)
Non-trainable params: 0 (0.00 Byte)
```

Fig. 4. Architecture of CNN model

E. Model training and testing

The model is trained underfitting over the training and validation dataset. Table 2 gives the information on every epoch. With each epoch, accuracy is consistently improved. Figure 5 shows the plot between Training vs Validation Accuracy and Training vs Validation loss. From the plot, training accuracy increased at regular intervals, but validation accuracy has some ups and downs.

Table II – Accuracy and loss for each epoch

Loss (for each epoch)	**Accuracy (for each epoch**	**Val_loss (for each epoch**	**Val_ Accuracy (for each epoch**
0.9518	0.6374	0.5534	0.8109
0.3167	0.8848	0.3414	0.8832
0.2520	0.9127	0.6211	0.7845
0.2019	0.9239	0.3772	0.8651
0.1490	0.9461	0.3716	0.8832
0.1797	0.9339	0.7346	0.7796
0.1572	0.9443	0.1255	0.9572
0.1426	0.9485	0.2090	0.9276
0.1255	0.9541	0.4312	0.8849
0.1215	0.9571	0.4479	0.8766
0.0938	0.9643	0.2068	0.9309
0.0821	0.9653	0.3154	0.8997
0.0770	0.9748	0.1762	0.9359
0.0954	0.9681	0.2742	0.9161
0.0837	0.9685	0.3876	0.8832
0.0814	0.9710	0.1503	0.9474
0.0899	0.9671	0.1042	0.9605
0.0941	0.9675	0.6306	0.8224
0.0856	0.9688	0.2319	0.9276
0.0603	0.9772	0.1458	0.9507

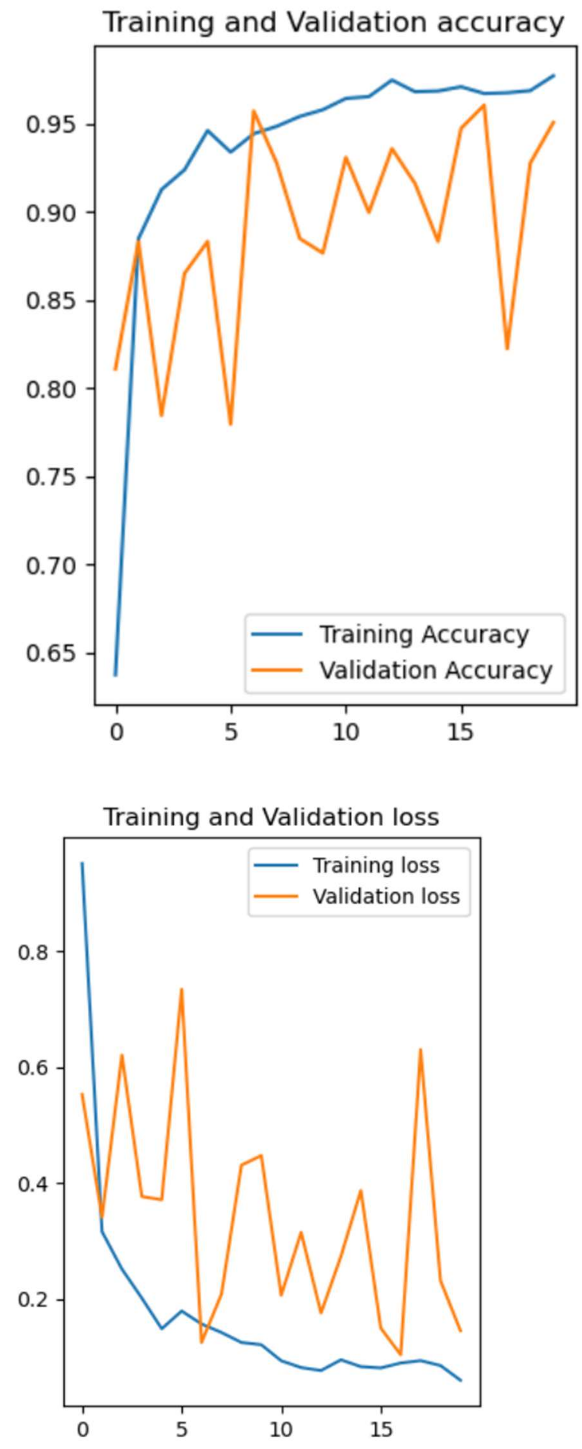

Fig. 5. Plots depicting Accuracy and Loss

IV. RESULT ANALYSIS

After the experimentation on training the model, it is tested and evaluated. On evaluation, it is observed that the accuracy is 96.5 on the testing dataset which is the best accuracy. Figure 6 shows the actual label and the predicted label by the model and the confidence value of the images.

Figure 5 shows the contrast of Training Loss and Validation loss and also Training accuracy and Validation accuracy. It is observed that using Tensorflow and Keras image classification is done with an accuracy of 95% which shows that the model has performed best on unseen data.

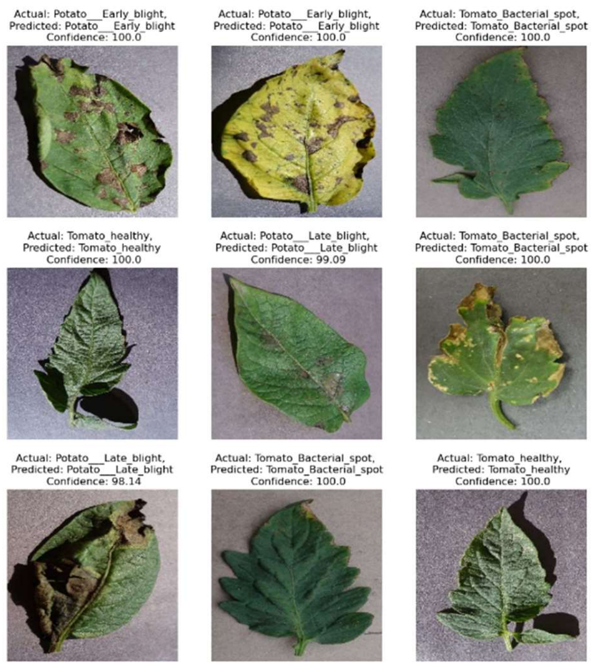

Fig. 6. Actual and predicted class of different images and confidence value

V. CONCLUSION AND FUTURE SCOPE

In conclusion, crop disease detection using image analysis has emerged as a promising and effective approach to safeguarding agricultural productivity and ensuring food security. The integration of image processing, computer vision, machine learning, and deep learning techniques has changed the way we recognize and combat plant diseases, offering significant advantages over traditional manual methods. This technology has the potential to transform farming practices by enabling early and accurate disease diagnosis, thereby minimizing yield losses and reducing the reliance on agrochemicals.

However, the most notable advancements have been achieved through the application of machine learning algorithms, especially deep learning approaches like Convolutional Neural Networks (CNNs). These neural networks have demonstrated remarkable capabilities in automatically learning relevant features from images, leading to highly accurate disease identification.

In conclusion, the integration of image analysis techniques with advanced technologies can transform crop disease management. By providing timely and accurate disease detection, this approach contributes significantly to sustainable agriculture, reducing economic losses and contributing to global food security. As research in this field continues to evolve, it is essential to promote knowledge dissemination, facilitate technology adoption, and ensure that the benefits of crop disease detection using image analysis reach farmers worldwide.

REFERENCES

[1] Thair A. Salih, Ahmed J. Ali, Mohammed N. Ahmed (2020) Deep Learning Convolution Neural Network to Detect and Classify Tomato Plant Leaf Diseases. Open Access Library Journal, 7,1-12, doi: 10.4256/oalib.1106296.

[2] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419.

[3] S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing," 2015 International Conference on Computing Communication Control and Automation, Pune, India, 2015, pp. 768-771, doi: 10.1109/ICCUBEA.2015.153.

[4] Jingcheng Zhang, Yanbo Huang, Ruiliang Pu, Pablo Gonzalez-Moreno, Lin Yuan, Kaihua Wu, Wenjiang Huang Monitoring plant diseases and pests through remote sensing technology: A review, October 2019, 104943.

[5] N. Kitpo and M. Inoue, "Early Rice Disease Detection and Position Mapping System using Drone and IoT Architecture," 2018 12th South East Asian Technical University Consortium (SEATUC), Yogyakarta, Indonesia, 2018, pp. 1-5, doi: 10.1109/SEATUC.2018.8788863.

[6] PlantVillage dataset. Available on

https://www.kaggle.com/datasets/emmarex/plantdisease