Leaf Disease Detection using Machine Learning Algorithms

Dr. G.Charles Babu[1]
GRIET, Hyderabad, Telangana
charlesbabu.griet@gmail.com

Syed Mizbahuddin[2]
Department of CSE
GRIET, Hyderabad, Telangana
saifsyed1000@gmail.com

Thouti Bharath Kumar[3]
Department of CSE
GRIET, Hyderabad, Telangana
thoutibharath@gmail.com

Sai Supreeth Goud Arukala[4]
Department of CSE
GRIET, Hyderabad, Telangana
supreethgoud9849@gmail.com

Naredla Phaneendra Reddy[5]
Department of CSE
GRIET, Hyderabad, Telangana
phaneendrareddynaredla@gmail.com

A . Shishir Kumar[6]
Department of CSE
GRIET,Hyderabad,Telangana
shishirkumar.akula@gmail.com

2023 2nd International Conference on Edge Computing and Applications (ICECAA) | 979-8-3503-4757-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICECAA58104.2023.10212425

Abstract- Plant diseases are mostly affecting leaves. In most of the cases, manual disease identification method fails to identify the disease correctly due to the similar symptoms of various diseases. People lack sufficient knowledge of plant diseases. The inability to detect the plant disease leads to crop production loss. Moreover, farmers have suffered significant losses as a result of a lack of sufficient understanding and direction to address the issue. This necessitates the need to develop a novel technology to detect the plant diseases. This study has attempted to develop an effective plant disease detection model using Convolutional Neural Networks (CNN). The proposed model has the ability to detect multiple diseases that occur in a single plant species. The results show the efficiency of the proposed model.

Keywords: Leaf disease, Convolutional Neural Networks, fertilizer

I. INTRODUCTION

Agriculture is the largest source of employment in India, particularly in rural regions. It also boosts the country's GDP. The main reason for lower crop output due to several leaf diseases is the farmer's inability to diagnose the plant diseases results in destroying hundreds of acres of crops every year. The agricultural sector is significantly impacted by various plant diseases.

To combat the diseases, a lot of fertilizers and pesticides are introduced every day. As old as plants themselves, plant diseases and insect damage have been caused, and various methods for preventing them have had variable degrees of effectiveness. In terms of enabling protection against certain diseases and pests, modern biotechnology techniques offer higher degree of protection. The efficacy of plant breeding is enhanced by the employment of both genetically modified plants and molecular mechanisms. Plant diseases could be treated by using a variety of techniques.

Before applying fertilizers to a crop, we must first determine whatever disease is present in order to determine which fertilizers should be utilized. The majority of issues will be resolved if the disease is detected in a plant.

This would make it easier for farmers to see plant diseases right where they are. They are no longer required to consult experts or scientists. The type of disease in the plant could be determined by analyzing at a single leaf. To produce accurate findings, we have utilized a variety of machine learning methods. In conclusion, we also offer a quick tip that aids in preventing and treating plant disease.

II. LITERATURE SURVEY

According to [1], integrating machine learning and image processing methods, namely Histogram of Oriented Gradient (HOG) for feature extraction and Random Forest for classification, may be used to identify whether a plant is healthy or sick. This study emphasizes that the developed approach can be particularly useful for farmers in rural areas who may not have easy access to agricultural experts.

In [2], researchers use the deep convolutional neural network (CNN) models to identify and diagnose plant diseases from images of leaves. To lower the number of parameters and processing costs, the authors substituted ordinary convolution with depth-separable convolution. The models outperformed standard handcrafted feature-based techniques in terms of classification accuracy for recognizing different plant illnesses.

The paper [3] emphasizes the importance of identifying and diagnosing plant diseases to improve food production worldwide. It demonstrates the use of deep learning models like InceptionV3, InceptionResnetV2, MobileNetV2, and EfficientNetB0 to detect plant diseases using images of healthy and diseased leaves. As compared to previous machine learning methodologies, the models exhibited excellent classification accuracy rates and needed less training time.

Overall, these publications show the possibility of employing deep learning and machine learning approaches to reliably and effectively identify plant diseases, which might have a huge influence on boosting food production and assisting farmers in rural regions.

III. METHODOLOGY

1. Convolutional Neural Network

Step1: Convolutional operation

The essential component of the proposed strategy is the convolution operation. At this phase, we'll look at feature detectors, which serve as filters for the neural network. In addition, we will discuss feature maps, the parameter learning process for these maps, pattern recognition, the many detection layers, and the mapping of found patterns.

Figure 1: Dataset

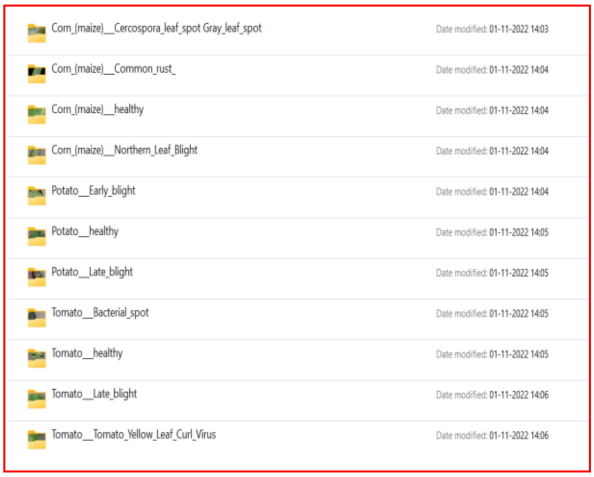

Figure 2: Diseases

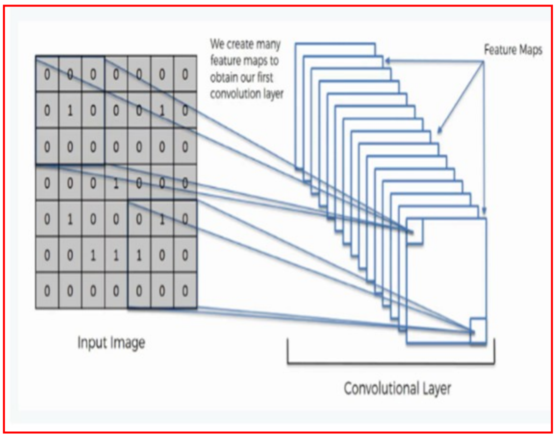

Figure 3: Input Image and Feature Detector

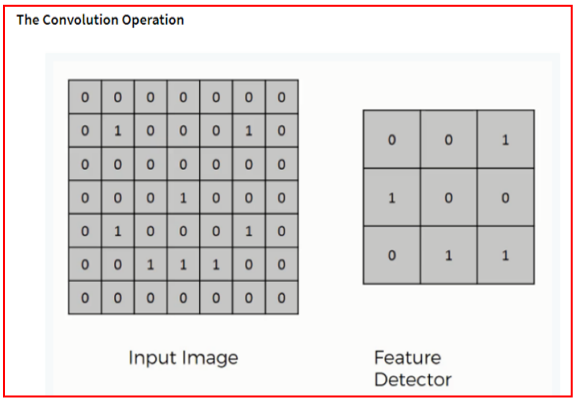

Figure 4: Convolutional Layer

Step (1b): ReLU Layer

In the following phase, we will concentrate on the Rectified Linear Unit (ReLU), which is an essential component of the process. he usage of ReLU helps to prevent the exponential growth in the computation required to operate the neural network. We will look at ReLU layers and how linearity works within the context of Convolutional Neural Networks. Although it is not required to understand CNNs, we may give a basic lesson to help you get started.

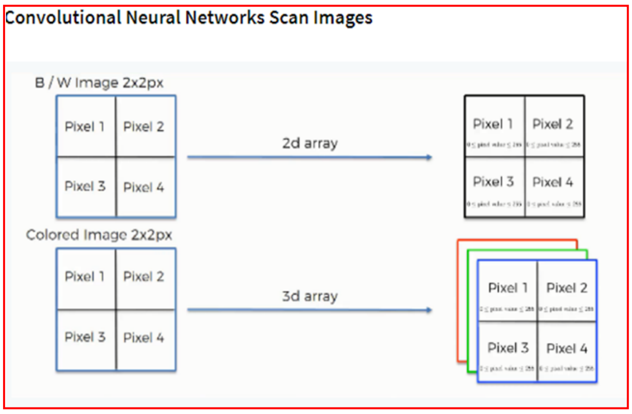

Figure 5: Image Scanning

Step 2: Conv2D

The Keras Conv2D layer is responsible for 2D convolution. A tensor of outputs is produced by this layer after a convolution kernel it generates is convolved with the input of the layer. For various image processing tasks including blurring, sharpening, embossing, edge recognition, and more, a kernel is a convolution matrix or mask. The kernel is convolved with an image to achieve this.

Step 3: Flattening

This layer is responsible for processing the data from pooling to flattened layers.

Step 4: Full Connection

This layers performs the major processes of Convolutional Neural Network. Here, the resultant "neurons" learn to identify the images.

IV. ALGORITHMS

Convolutional neural network (CNN):

There are three layers that make up the architecture of a Convolutional neural network: input, hidden, and output. concealed layers are any intermediate layers in a feed-forward neural network that have their inputs and outputs concealed by the activation function and final convolution.

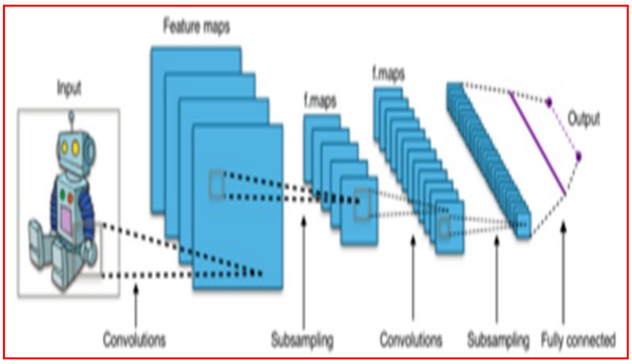

Figure 6: CNN Architecture

Artificial Neural Network (ANN):

Artificial Neural Networks (ANN) was developed as a result of research into the structure and operation of human brain networks. Similar to brain neurons, ANN is composed of layers of neurons. The feed-forward neural network is one common form of ANN. It contains three layers: an output layer that provides the problem solution, an input layer for receiving external data to conduct pattern recognition, and a hidden layer that serves as a bridge between the other levels. Neurons in the input layer and output layer are connected by acyclic arcs. The ANN uses a training strategy that modifies the neuron weights based on the error rate between the desired and actual output to learn from the datasets. In most cases, the back propagation method is utilized as the training procedure in ANN. The diagram below depicts the overall structure of an ANN.

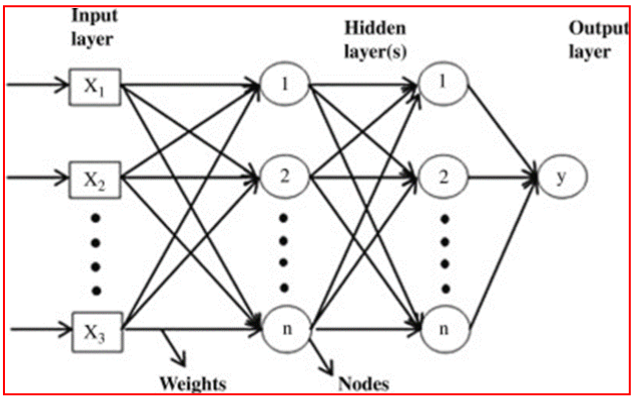

Figure 7: ANN

ResNet50:

The ResNet50 Convolutional Neural Network (CNN) has 50 layers. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun made and prepared it in 2015, and the model presentation results might be found in their examination article named "Profound Remaining Learning for Picture Acknowledgment." Similar to VGG-19, this model has been trained on more than one million images from the Image Net collection and can identify up to one thousand items. Colored 224x224-pixel images served as training grounds for the network. ResNet50 is a large-scale model with accurate image recognition tasks in terms of size and performance.

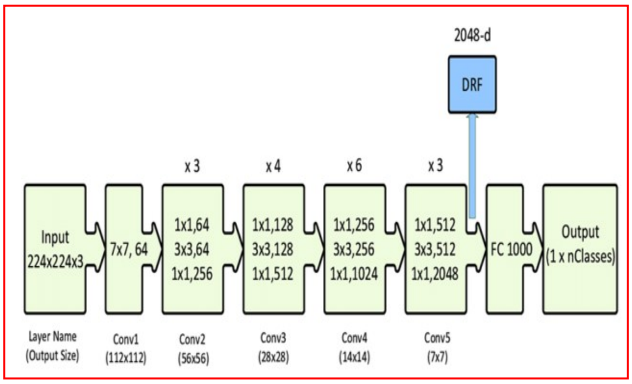

Figure 8: ResNet 50 Architecture

Finally, ResNet has significantly influenced how deep convolutional neural networks are trained for computer vision applications. ResNet50, a newer variant, uses 3-layer bottleneck blocks to increase accuracy while reducing training time, whereas the original ResNet had 34 layers and used 2-layer blocks. Keras, a popular deep learning API, has various pre-trained models, like ResNet50, that are easily available for testing. Creating a ResNet model for image classification in Keras is simple and can be completed in a few simple steps.

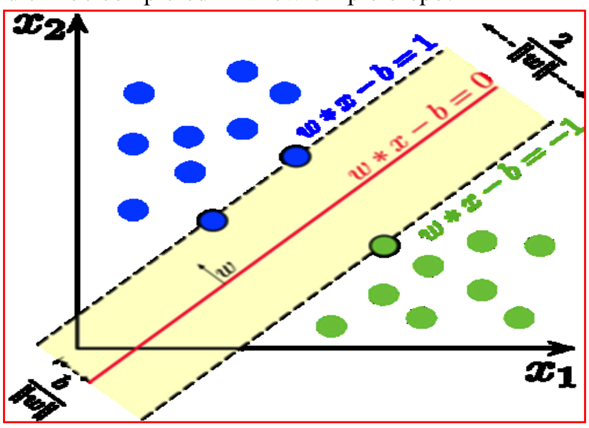

Figure 9: SVM

Support vector machine (SVM):

Despite the fact that the Support Vector Machines (SVM) may be used for handling both classification and regression problems and processing a wide range of continuous and categorical data, they are typically categorized as a classification approach. In a multidimensional space, SVM produces a hyperplane to distinguish between multiple classes. In order to reduce error, it builds ideal hyperplanes repeatedly. Finding the Maximum Marginal Hyperplane (MMH) divides the dataset into distinct groups.

Support Vectors

In Support Vector Machine (SVM), the support vectors are the particular data points that are closest to the Hyperplane. The Hyperplane may be defined more precisely by employing the support vectors to calculate margins, resulting in a strong classifier. As a result, the support vectors are critical in the development of the SVM classifier.

Hyperplane

A hyperplane is a plane that is used to divide a collection of objects into those that belong to distinct classes.

Margin

The margin is the distance between the two lines drawn closest to them. The nearest points or support vectors closest to the line are measured and the perpendicular distance between them is calculated. A great margin is the one with a larger spread, whereas a bad margin is the one with a lower spread.

The basic goal is to split the provided dataset precisely. Hence, SVM selects the Hyperplane with the best edge between the support vectors in the dataset. To select the Hyperplane with the greatest margin, the following steps should be performed:

Build a Hyperplane that divides the classes properly. Three Hyperplanes are depicted on the left-hand side image: black, blue, and orange. The black one successfully divides the two classes, whereas the blue and orange ones have more classification errors.

As indicated in the right-hand image, choose the hyperplane with the maximum segregation from the closest data points.

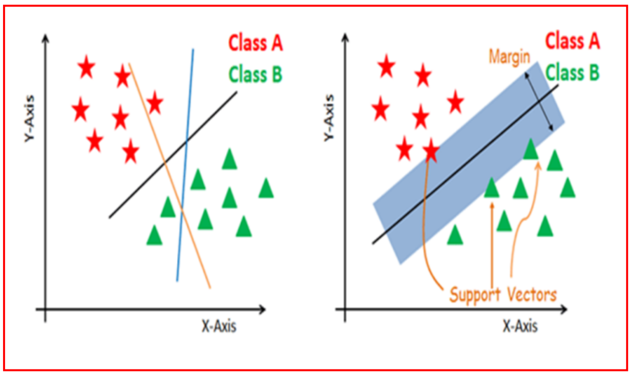

Figure 10: Kernel

SVM Kernels

In reality, a kernel is often used to implement the SVM approach. The kernel is responsible for transforming the input data into the required format. The kernel method is a SVM approach that transforms a low-dimensional input feature into a higher-dimensional feature. This strategy is very beneficial for handling nonlinear separation issues and emerges as a more accurate classifier.

The ability of SVM to handle classification and regression issues is its primary advantage. In order to divide or classify any two classes, it does this by building a decision boundary or Hyper plane between them. SVM is also used for performing object identification and image classification.

Polynomial Kernel: The polynomial kernel is an extension of the linear kernel that can recognize curved or non-linear input spaces.

V. PROCEDURE OF USAGE

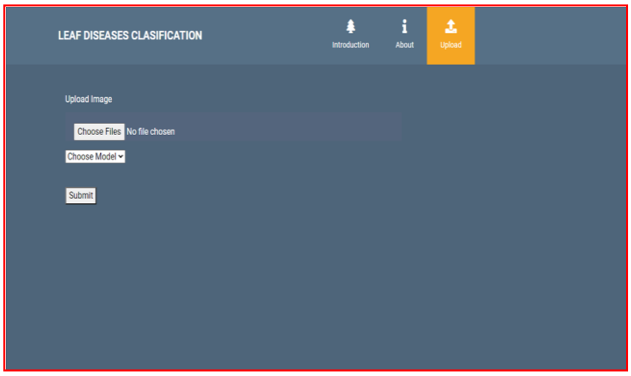

Figure 11: Home

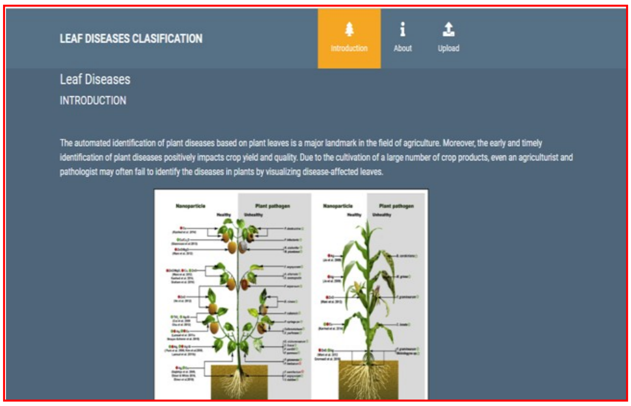

Figure 12: About

About: Distinguishing between various diseases and the healthy state of leaves

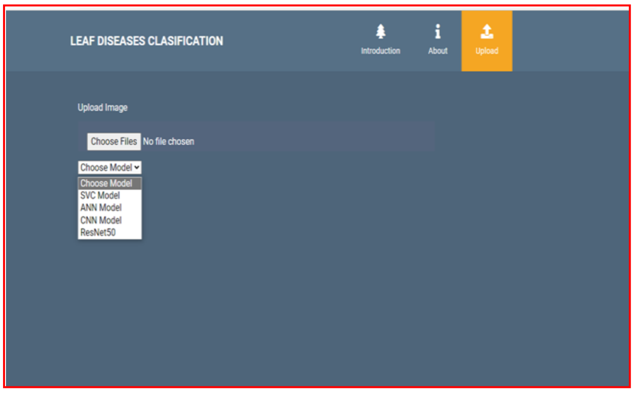

Figure 13: Image Uploading

Image upload: Select the test image from your file system

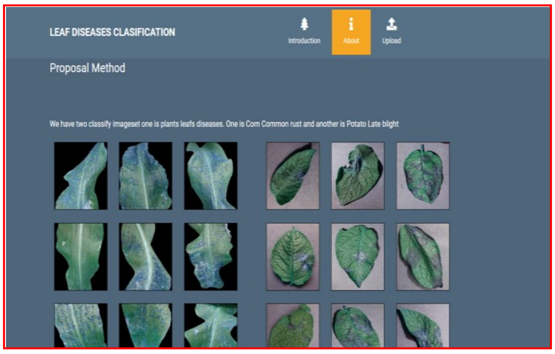

Figure 14: Model choosing

Model choosing: Choose one out of the 4 available models.

VI. RESULTS AND ANALYSIS

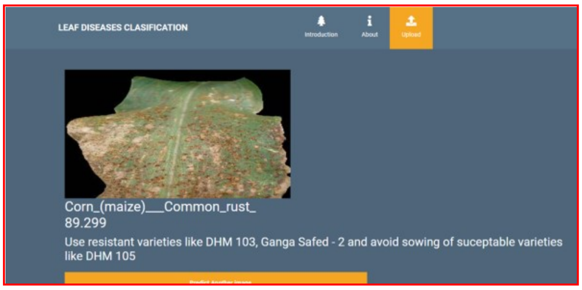

Figure 15: Classified output

Classified output:

The uploaded image is labelled Corn_(maize)___healthy.

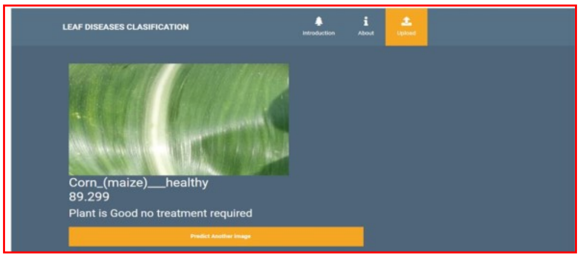

Figure 16: Classified output

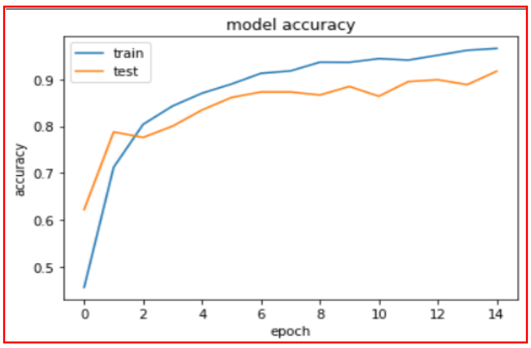

Figure 17: CNN

The uploaded image is labelled Potato___Early_blight.

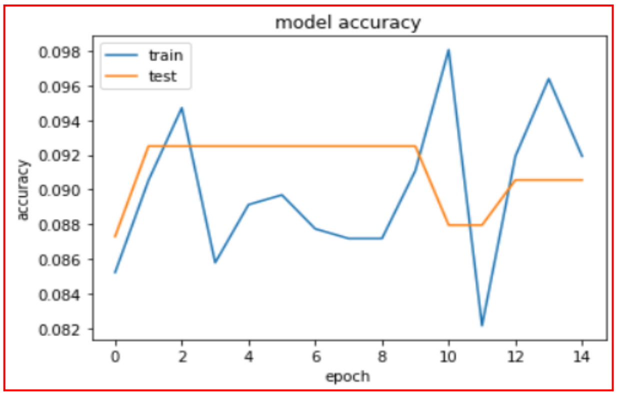

Figure 18: ANN

The Support Vector Machine (SVM) model was the most accurate with a score of 0.965. This shows that, SVM is good at identifying and detecting plant diseases from leaf images [1]. SVM is a machine learning technique that is well-known for its ability to process high-dimensional input image. It operates by locating a hyper plane that divides the classes in the feature space.

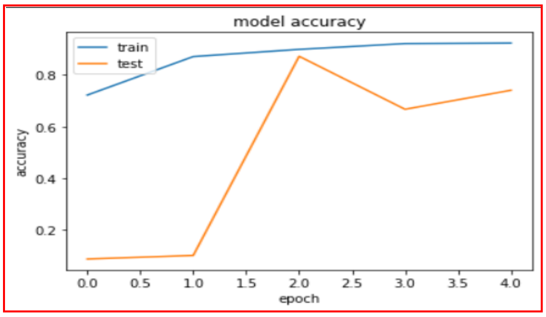

Figure 19: SVM

The Convolutional Neural Network (CNN) algorithm achieved an accuracy of 0.8929. CNNs have been widely used for performing image classification tasks and have outperformed traditional feature-based approaches [2]. The CNN architecture consists of convolutional layers, pooling layers, and fully connected layers to automatically identify different characteristics from input images.

The ResNet50 algorithm attained a high accuracy of 0.92344. ResNet is a deep neural network architecture created to address the issue of gradients in deep neural networks. It uses residual networks to achieve an improved information flow and more efficient training. ResNet has been demonstrated to deliver cutting-edge performance on a variety of image classification tasks [3].

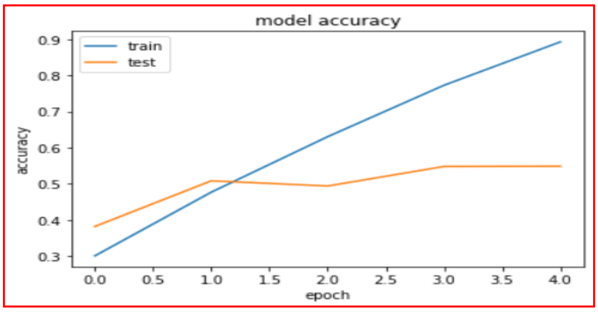

Figure 20: ResNet50

The accuracy of the Artificial Neural Network (ANN) technique was 0.0919. This is not unexpected given that ANN is a very basic neural network design that is unsuitable for image classification applications. ANN is generally made up of three layers: an input layer, one or more hidden layers, and an output layer. It is frequently used for basic classification challenges with low-dimensional input characteristics.

```
model    : accurary

SVC      : 0.9659122824668884
ANN      : 0.09192512184381485
CNN      : 0.8929868936538696
ResNet50 : 0.9234423041343689
```

Figure 21: Accuracy

VII. CONCLUSION & FUTURE SCOPE

Using machine learning approaches, the proposed model has effectively identified the plant diseases. Further the obtained images are classified by using 11 distinct types of plant diseases as our training data. Additionally, our approach can identify healthy plants. We trained using a variety of models, including CNN, ANN, SVC, and ResNet50, on a sizable dataset from the Kaggle database. The scope of this study might be expanded to include the detection of plant diseases. This can be further developed to offer recommendations for detecting and analyzing various diseases based on the soil type. This research work can be expanded to identify more plant diseases.

REFERENCES

[1] A. Camargo and J.S. Smith, "An image-processing based algorithm to automatically identify plant disease visual symptoms," Biosystems Engineering, vol.102, pp.9–21, January 2009.

[2] J.S. Cope, D. Corney, J.Y. Clark, P. Remagnino, and P. Wilkin, "Plant species identification using digital morphometrics: A review," Expert Systems with Applications, vol.39, pp.7562–7573, June 2012.

[3] J. Garcia and A. Barbedo, "Using digital image processing for counting whiteflies on soybean leaves," Journal of Asia-Pacific Entomology, vol.17, pp.685–694, December 2014.

[4] A. Gongal, S. Amatya, M. Karkee, Q. Zhang, and K. Lewis, "Sensors and systems for fruit detection and localization: A review," Computers and Electronics in Agriculture, vol.116, pp.8–19, August 2015.

[5] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

[6] J. Hemming and T. Rath, "Computer-vision-based weed identification under field conditions using controlled lighting," Journal of Agricultural Engineering Research, vol.78, no.3, pp.233–243, March 2001

[7] S. Ji-Yong, Z. Xiao-Bo, Z. Jie-Wen, W. Kai-Liang, C. Zheng-Wei et al., "Nondestructive diagnostic of nitrogen deficiency by cucumber leaf chlorophyll distribution map based on near infrared hyperspectral imaging," Scientia Horticulturae, vol.138, pp.190–197, May 2012.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Proceedings of 25th Advances in Neural Information Processing Systems (NIPS 2012), pp. 1097–1105,2012.

[9] C. Leksomboon, Plant Disease and Diagnosis, Kasetsart University Press, 2011 (in Thai).

[10] P. Li, S.H. Lee, and H.Y. Hsu "Review on fruit harvesting method for the potential use of automatic fruit harvesting systems," Procedia Engineering, vol.23, pp.351–366, 2011.

[11] T. Liu, W. Chen, W. Wu, C. Sun, W. Guo, and X. Zhu, "Detection of aphids in wheat fields using a computer vision technique," Biosystems Engineering, vol. 141, pp.82–93, January 2016.

[12] P. Noinongyao, U. Watchareeruetai, P. Khantiviriya, C. Wattanapat Boonsuk, and S. Duangsrisai, "Separation of abnormal regions on black gram leaves using image analysis," Proceedings of the 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE 2017), 2017.

[13] L.M. Romualdo, P.H.C. Luz, F.F.S. Devecchio, M.A. Marin, A.M.G. Zu´niga, O.M. Bruno, and V.R. Herling, "Use of artificial vision techniques for diagnostic of nitrogen nutritional status in maize plants," Computers and Electronics in Agriculture, vol.104, pp.63–70, June 2014.

[14] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," Computers and Electronics in Agriculture, vol.72, pp. 1–13, June 2010.

[15] Y. Song, C.A. Glasbey, G.W. Horgan, G. Polder, J.A., Deleman, G.W.A.M. van der Heijden, "Automatic fruit recognition and counting from multiple images," Biosystems Engineering, vol.118, pp.203–215, February 2014.