**PLANT LEAF DISEASE DETECTION USING MACHINE LEARNING**

**A COMPREHENSIVE REPORT**

**GITHUB LINK:** [**https://github.com/vishnu43/NN\_Project**](https://github.com/vishnu43/NN_Project)

**Abstract:**

Agriculture’s productiveness is something on which the economy hangs. Climate and other environmental conditions cannot be administered by farmers. Disease reduction is a key factor to consider in the context of agriculture. Crops affected by pests or diseases should be considered immediately. If proper leaf care is done at the outset, it can prevent the spread of disease to plants. On the alert, with hard eyes, detecting the disease with a leaf is difficult. This can lead to improper use of pesticides and ultimately lead to crop failure. Due to several environmental factors, several diseases affect plants. This leads to reduced crop quality and production. An automated disease-detection system is mandatory as this system will be useful for monitoring plants which is why immediate steps can be taken. It has been found that imaging techniques provide effective results in diagnosing diseases. Here a machine-based program helps farmers identify plant diseases by inserting a leaf image into the system. The program contains a set of algorithms for pre-screening, feature extraction, and based machine programming that can identify the type of disease.

**Keywords:**

Machine Learning, Neural Networks, Convolutional Neural Networks, Hue Saturation Value, Image Processing, TensorFlow, Keras, Data preprocessing, Standardization, Classification, Predictive System

**1. Introduction**

**1.1 Background**

Diseases can be caused by bacteria or any virus also it may be a fungal one that inflicts diseases that cause the loss of money every year, thus affecting the country and its people. Most of the time, the diseases can be identified through the leaves or by the trunk of the tree, sometimes also by its root, but leaves are easy to identify among others. Because of the difficulty in checking each plant root, the disease and its causes are mostly ignored by the farmers. Hence, Diseases are detected on the leaves of the plant. Therefore, Technology is needed to help the farmers identify the disease and take preventive measures such that it helps recover the crop that is to get infected in the future. Farmers can avoid going to the labs as this technology helps them identify the disease on their own. This type of Detection might prove gain in looking at huge fields of crops, and leaf disease detection analysis helps in identifying the diseases.

**1.2 Objectives**

Design efficient CNN models for plant disease detection with reduced memory and computational requirements, optimized for deployment on mobile and IoT platforms with limited hardware. Develop depth-wise separable convolution architectures that strike a balance between model size, performance, and processing speed. Compare the performance of these new models with standard CNN architectures like MobileNet, VGG16, and AlexNet.

**2. Motivation:**

Agriculture is the backbone of many countries, with small farmers contributing nearly 80% of total production in developing regions. However, crops are often vulnerable to external factors that can lead to infections, resulting in significant losses. Many primary symptoms of plant diseases are microscopic, limiting the human ability to detect them visually and making traditional inspection methods both time-intensive and challenging. This highlights the need for an automated system capable of detecting, classifying, and quantifying disease symptoms efficiently. A variety of diseases typically affects the leaves, fruits, and stems of plants, and a computer vision system could reduce human involvement while improving accuracy in identifying and classifying crop diseases. Such automated detection would be particularly beneficial in large fields, enabling early detection across extensive areas. Image processing plays a major role in enabling these systems to recognize disease symptoms effectively.

**3. Main Contributions & Objectives:**

* Proposed depthwise separable convolution models, such as customized MobileNet variants, specifically designed for detecting plant diseases
* Compiled a new dataset of 6,580 plant leaf images for rigorous testing and evaluation of the models.
* Showcased significant improvements in computational efficiency, memory usage, and classification accuracy.
* Achieved high disease detection accuracy (up to 99.55%) using the developed models.
* Conducted thorough comparisons with traditional CNN architectures like VGG16, and AlexNet, demonstrating the new models' advantages.

**4. Proposed Framework:**

Proposed a machine learning application based on Convolutional Neural Networks (CNN) to help farmers identify plant diseases by uploading leaf images. The model employs an algorithm for feature extraction and classification, which can accurately identify the disease type and suggest remedies. Images in the dataset are resized to maintain a consistent filter size, optimizing computational efficiency and aligning with the model architecture for faster, more accurate results. For feature extraction, we utilized color co-occurrence to highlight affected areas on leaves. The CNN-based deep learning algorithm processes images by breaking them down into arrays without losing essential features, enabling accurate predictions. Through dimensionality reduction, the images are simplified while retaining necessary details for effective disease identification.

**5.Data Description:**

The input given to the model are images from the plant village dataset comprised of details of plant leaves with the diseases accordingly. The types of plant leaves used in this are tomato, potato, and pepper leaves respectively.

**6. Design/Framework**

**6.1 Dataset Overview**

The Plant Village dataset, comprising various pictures of leaves features, was chosen for its comprehensive nature. A detailed exploration of the dataset's characteristics and distributions laid the groundwork for subsequent processing steps.

**6.2 Data Processing**

The data set consists of images, so noise removal was not an important task here in this process. The images in the dataset were resized to make the model training easier to develop. The technique here of input helps speed up the training process. The images of the leaves are after preprocessing as they are in one direction that is upwards. Here, they all are rotated and arranged in one way.

**6.3 Image Segmentation**

Its role in image processing is to divide an input image into distinct regions or parts based on certain characteristics, making the process simpler. There are many methods available for this, such as boundary-based methods and spot detection algorithms, among others.

**6.4 Convolution Neural Networks**

CNN (Convolutional Neural Networks) are deep neural networks where neurons are interconnected to create a receptive field for the input image. The goal is to reduce image dimensions using filters, preserving essential features while minimizing pixel size. Popular CNN architectures, such as AlexNet, GoogLeNet, and VGGNet, are widely used. This approach has significantly influenced engineering research, with applications like identifying damaged parts of leaves in agriculture.

**7. Implementation/Methodology**

**7.1 Train-Test Split**

Training and testing the data are crucial steps in building a machine-learning model. The dataset is typically split into two parts: the training set and the testing set. Using the same data for both training and testing can lead to overfitting and discrepancies, negatively impacting the model's accuracy and efficiency. The dataset should be divided in such a way that the testing set is sufficiently large. Generally, the testing set is held out from training and later used for cross-validation, helping assess the model's accuracy.

**7.2 Feature Extraction**

Feature extraction can primarily focus on shape-based, color-based, and venation-based methods. Shape-based extraction involves aspects like aspect ratio, area, and rectangularity, which vary between leaves. Venation-based extraction uses information from vein patterns. In this project, we focus on color-based extraction, which utilizes the color co-occurrence technique. This method extracts feature points based on the geometric center of the image, converting it into color blocks. The analysis is done using HSV (Hue, Saturation, Value) numerical readouts.

**7.3 Model Training**

The convolutional model consists of five key layers, starting with an input layer, where a pre-processed image is fed into the model. It then passes through the Conv2D layer, followed by a MaxPooling layer. The connected neurons are finally combined in the fully connected layer, leading to the output layer, which projects the final vector or output. To identify the mitigated part of a leaf, images are used as input. This architecture allows us to extract features from the images. By leveraging the complex features of the image, the Conv2D and MaxPooling layers can be further enhanced to capture more detailed information.

**7.4 Model Evaluation**

The fully connected layer of neurons at this stage takes input from the outputs of previous layers. This information is then passed to the softmax layer, which produces the final output, typically a vector or image, that helps identify the disease. Based on this output, we can devise a remedy to address the problem, providing a solution for managing the disease in the field.

**7.5 Predictive System**

The prediction of the mitigated part of the leaf is made by processing the input image through several layers to extract features. If the image is complex, the pooling and convolution layers are increased to capture more detailed information. The fully connected layers then take the outputs from the previous layers and combine them into a single vector for the final classification, identifying the mitigated part of the leaf. This process enables disease prediction by analyzing the image and detecting the affected areas.

**8. Testing/Critical Assessment**

The use of depthwise separable convolutions reduces processing requirements, enabling compatibility with mobile and IoT devices. An accuracy of 99.55% highlights the effectiveness of the model in practical settings. It effectively addresses a range of plant diseases across different species, showcasing its adaptability. The complexity of deep learning models may impact their decision-making processes, making it harder for agricultural professionals to interpret results. The existing dataset of 6,580 images could benefit from increased size and diversity for better representativeness. Further evaluation in diverse environmental contexts may be necessary to validate the model's robustness**.**

**9. Deployment/Recommendations**

**9.1 Deployment Considerations**

The color co-occurrence method is used to detect color changes in leaves, helping determine whether a leaf is healthy or diseased. This method works on the HSV color space derived from RGB images. It can detect diseases like late blight, bacterial infections, and other viral diseases on tomato and potato leaves, which are used as input in this project. By analyzing the images, we can identify the affected areas and recommend targeted pesticide or insecticide applications to prevent the rapid spread of these diseases across the crop.

**9.2 Recommendations for Improvement**

Future directions for model improvement are discussed, including accessibility for all kinds of farmers through IoT & Mobile devices, exploring alternative neural network architectures, and incorporating additional relevant features from emerging research.

**10. Results/Experimentation & Comparison/Analysis:**

The models' effectiveness was rigorously analyzed using various metrics, such as accuracy, F1-score, computational efficiency, and memory utilization. Their results were compared against established CNN architectures like VGG16, and AlexNet to assess the new models' effectiveness. This evaluation not only underscored the depthwise separable convolution models' high accuracy but also showcased their resource efficiency. The comparison aimed to demonstrate their practicality for real-world agricultural applications

**11. Conclusion**

The proposed system classifies diseases in potato and tomato leaves using a Convolutional Neural Network (CNN). It was implemented in Python using Google Colab, with input images acquired from publicly available datasets. The system uses color co-occurrence techniques to detect diseases such as late blight, bacterial, and viral infections on the leaves. Once the affected areas are identified, the required pesticide can be applied to prevent the spread of disease across the crop, thus protecting the entire harvest and benefiting public health. The model's accuracy is lower compared to other projects because it classifies images of three different plants, whereas other models typically work with a single plant species. Two conclusions were drawn: the first focuses on identifying the plants, and the second on diagnosing their diseases.

**12. References**

• [1] Qiufeng Wu, Yiping Chen, Jun Meng, “DCGAN-Based Data Augmentation for Tomato leaf Disease Identification.” IEEE 2020

.•[2] Jianyu Su, Meihua Wang, Zhenxin wu, Qingliang Chen, “Fast Plant Leaf Recognition using Improved Multiscale Triangle Representation and KNN for Optimization”, IEEE 2020.

• [3] Bhumika S.Prajapati, Vipul K.Dabhi Harshadkumar, B.Prajapati, “A Survey on Detection and Classification of Cotton Leaf Diseases”, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016.

• [4] Sachin D. Khirade, A.B Patil ,“Plant Disease Detection Using Image Processing”, International Conference on Computing Communication Control and Automation”, 2015.

• [5] Malvika Ranjan, Manasi Rajiv Weginwar, NehaJoshi, Prof.A.B. Ingole, “Detection and classification of leaf disease using artificial neural network”, International Journal of Technical Research and Applications, 2015.

**13. Figures and Tables**

**A diagram of different types of boxes

Description automatically generated**

**A table with numbers and a text

Description automatically generated with medium confidence**

**A table of data with text

Description automatically generated with medium confidence**

**A screenshot of a test

Description automatically generated**

**A table with numbers and text

Description automatically generated**

.

.

A diagram of a diagram

Description automatically generatedA diagram of a data processing process

Description automatically generated