IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PLANTS LEAF DISEASE DETECTION USING MACHINE LEARNING ALGORITHMS

[1]Prof. Mr. A. A. Hipparkar, [2]Mr. Pratik Gaikwad, [3]Mr. Prajwal Kanawade, [4]Mr. Vitthal Dhade, [5]Mr. Kiran Ghodake

[1]Head Of Department, [2]Research Scholar, [3] Research Scholar, [4] Research Scholar, [5] Research Scholar

[1]Department of Computer Engineering ,

[1]PES's College of Engineering, Phaltan, Maharashtra, India

Abstract: This paper explores the efficacy of utilizing a pre-trained ResNet34 model for agricultural disease identification via image recognition. With a focus on a prevalent vegetable, the study seeks to raise farmers' awareness of innovative technologies to combat plant diseases. Employing a machine learning and image processing approach, an accurate algorithm is proposed for detecting leaf diseases in vegetables and plants. The investigation centers on samples of plant leaves exhibiting disorders, aiming to assist farmers in early disease detection. Leveraging the widespread availability of smartphones and advancements in computer vision, particularly convolutional neural networks (CNNs), the research aims to address the persistent threat of plant diseases to smallholder farmers' livelihoods and food security.

The developed model demonstrates the ability to distinguish between healthy leaf tissue and seven distinct plant illnesses. Deployed as a user-friendly web application, the model ensures accessibility for farmers. Training and validation utilize a meticulously curated dataset of leaf images captured under controlled conditions. The proposed methodology achieves impressive results, with an accuracy rate of 97.2% and an F1 score exceeding 96.5% based on rigorous validation procedures.

Amid rapid population growth, agriculture plays a critical role in sustaining human life. Early disease prediction is essential for food security, yet remains challenging. This paper aims to equip farmers with knowledge of cutting-edge technologies to mitigate plant diseases, particularly focusing on a widely available vegetable. Through the proposed machine learning and image processing approach, coupled with advanced algorithms, farmers can enhance disease detection in vegetables and plants, thereby safeguarding agricultural yields and ensuring food security.

***Keywords* –** CNN, Deep Learning, ResNet, Transfer Learning, Classification, Plant Disease Detection.

I. INTRODUCTION

In this paper, we investigate plant diseases, i.e., diseases affecting plants, using machine learning algorithms to identify the types of diseases and determine which remedies might be effective. Upon identifying the type of disease, we automatically suggest suitable treatments to farmers through a completely online process, leveraging models trained to provide such recommendations. Our endeavor is to provide farmers with these remedies seamlessly by conducting research and efforts in this direction.

Modern technologies have significantly boosted food production to meet societal demands, yet ensuring the safety and security of food crops remains a persistent challenge. Factors such as climate change, declining pollinator populations, and plant diseases continue to threaten farmers. Manual detection of plant diseases via human visualization is difficult, inefficient, and often relies on a limited set of leaf images, leading to timeconsuming processes. Alternatively, automatic identification techniques offer quicker and more accurate solutions, such as using image processing to detect diseases.

Addressing these challenges is crucial, and leveraging current technologies for analysis and detection processes can assist farmers in mitigating these issues. Plant diseases not only jeopardize crop yields but also pose risks to human life by potentially causing droughts and famines, resulting in significant losses, especially in commercial farming ventures. Technologies like computer vision and machine learning (ML) emerge as powerful tools in combating these diseases. In this paper, we propose an ML-based solution for identifying and managing plant diseases, comprising three key stages: identification, analysis, and verification using available databases. By integrating ML into these processes, we aim to provide farmers with effective tools for disease management and crop protection.

Our work focuses on developing a leaf recognition technique based on specific characteristics derived from images. Images are classified into four categories based on the code names of plant diseases. During the training phase, a variety of images are loaded and resized, and their corresponding labels are appended to the list.

For classification, we employ the CNN algorithm, consisting of several layers for efficient implementation, including convolutional layers, pooling, and a regression layer for output. Another critical parameter is the learning rate (LR), which determines the speed at which the model learns. After building the model, data is loaded, and a variable representing the model name is used to save the model in a folder. Subsequently, data is inputted into the model for detection.

Common plant diseases, such as viral, bacterial, and fungal infections, pose challenges in identification due to varying symptoms. Smaller datasets can hamper model performance, while larger ones can improve accuracy by mitigating overfitting. The quality of training data significantly influences model capabilities, with noise affecting classifier accuracy. Early detection is crucial due to limited datasets, and this system utilizes existing images and datasets for disease identification, offering a cost-effective approach. It employs CNN for detecting leaf health and identifying diseases like fungi, viruses, bacteria, etc., while also suggesting remedies for recovery.

II. Problem Statement

The significance of agriculture to the Indian economy cannot be overstated, but its commercialization has led to adverse environmental effects. Increased use of chemical pesticides has resulted in chemical accumulation in various environmental components, including soil, water, air, animals, and humans. Artificial fertilizers, while providing short-term yield boosts, have long-term detrimental effects on the environment, persisting for years and contaminating groundwater.

This trend has also negatively impacted farming communities worldwide, with farmers experiencing declining fortunes despite purported increases in productivity. The detection of plant diseases through visual observation of symptoms on leaves is challenging due to the vast array of cultivated crops and associated pathologies. Even experienced agriculturalists and plant experts may struggle to accurately identify specific diseases, leading to erroneous conclusions and ineffective solutions. Various fungal and bacterial diseases affect numerous plants, exacerbated by factors such as population growth and environmental contamination of air, water, and soil.

III. Literature Revie

Over time, significant research efforts have been dedicated to detecting leaf diseases through image processing, with continued interest driving ongoing investigations in this domain. Automatic crop disease detection, facilitated by image processing and machine learning, has emerged as a prominent area of study in recent years. This approach offers efficient analysis of large datasets, providing farmers with timely insights to effectively manage crop diseases. As interest and technological advancements persist, there is considerable potential for further developments in this field, promising improved agricultural practices and enhanced crop health.

According to the study conducted by Chowdhury et al. [1], tomato leaf images were classified for tomato diseases using the Efficient Net CNN architecture [2]. The Plant Village dataset, containing 18,162 tomato images, was used to fine-tune and train the model to detect healthy and diseased tomato leaf images. In the preprocessing step, they down sized images to 224 × 224. The results indicate that the proposed model outperformed various contemporary DL algorithms.

Cruz et al. [3] implemented a CNN network based on the LeNet architecture for detecting olive quick decline symptoms.

They initially trained the network on the Plant Village dataset, then retrained on a custom dataset. They improved the performance of the network by providing information about the edge patterns and the shape information. In the preprocessing step, they downsized images to 256×256. The authors reported an accuracy of 99%.

P. Krithika et al., [6] pre-processed by image resizing, contrast enhancement and color-space conversion. The K-Means clustering for seg- mentation and feature extraction using GLCM is performed. Classification was made using multiclass SVM. R. Meena et al., [7] performed color space conversion followed by enhancement process. The primary colors of leaves are converted into L ∗ A ∗ B ∗ . The K-Mean clustering algorithm is used for segmentation.

K. Muthu Kannan and co-workers located spot infections in leaves and classified them in line with the diseased leaf classes the usage of numerous devices studying algorithms. LVQ-Learning Vector Quantization, FFNN - Feed Forward Neural Network, and RBFN Radial Basis Function Networks were utilized to diagnose diseased plant leaves by analysing the collection Of shape and texture statistics from the troubled leaf picture. The simulation confirmed that the proposed machine is effective. With the support of this work, a machine learning-based system for improving crop quality in the Indian economy can be developed.[10].

IV. Methodology

4.1 Convolutional-Neural Network Model

A Convolutional Neural Network (CNN) is a deep learning model commonly used in machine learning for tasks involving image classification, object detection, and image segmentation, among others. CNNs are designed to automatically and adaptively learn hierarchical patterns and features from raw input data. In contrast to classic fully connected neural networks, Convolutional Neural Networks (CNNs) are designed for tasks like image processing. They incorporate specialized layers such as convolutional and pooling layers, enabling them to efficiently extract hierarchical features from structured data like images. These layers allow CNNs to effectively capture spatial relationships, making them highly effective for tasks such as image classification and object detection.

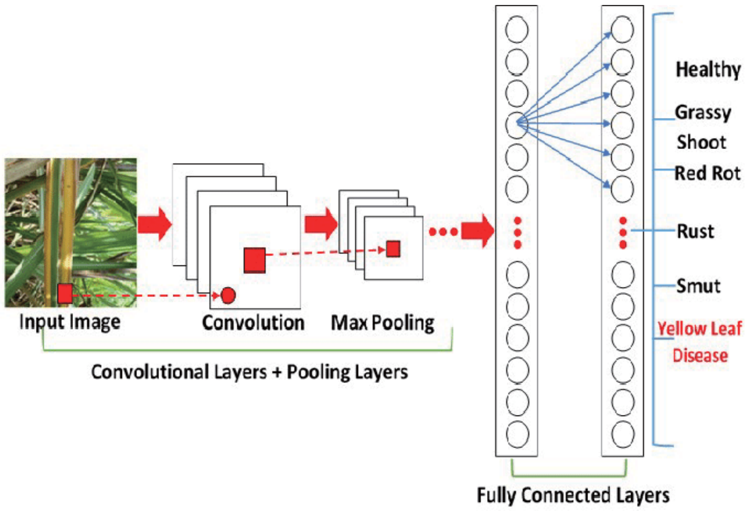

Fig 4.1: Convolutional-Neural Network Architecture

Here's a basic overview of the CNN architecture:

1. **Convolutional Layers**:
   These layers apply convolution operations to input images. Each convolution operation involves sliding a small filter (also known as a kernel) over the input image, computing the dot product between the filter and the overlapping regions of the image. This process extracts features from the input images.

2. **Input Layer**:
   It contains data in the form of image. The parameters include height, width, depth and color information of the image (RGB). Input size is fixed to 224 X 224 RGB image.

3. **Pooling Layer**:
   The pooling layer helps to reduce the computational power in order to process the data by decreasing decreasing (or) reducing the dimensions of the featured matrix obtained by using the dot products. Fully connected layer: It comprises of loads, neurons and biases. It connects neurons from one convolutional layer to another.

4. **SoftMax Layer/ Logistic Layer**:
   1. **SoftMax** is typically used for multi-class classification tasks. It takes a vector of arbitrary real-valued scores and converts them into probabilities that sum up to 1. Each score is transformed into a probability indicating the likelihood of the input belonging to each class.

   2. **Logistic Regression**: Logistic regression, often used as the activation function in the output layer for binary classification tasks, calculates the probability that a given input belongs to one of the two classes (usually denoted as 0 or 1). It squashes the output of a linear function between 0 and 1, representing the probability of the positive class.

5. **Activation Function**:
   ReLU: It transforms the total weighted input through the node and puts it into the operation, activates the node. Rectified Linear Unit (ReLU) is an activation function used in the neural networks for convolutional
   Operations.

6. **Fully Connected Layer:**
   After several layers of convergence and concentration, CNN usually ends with several fully connected layers. The multidimensional arrays (tensor) that have at the output of these layers are converted to vector and then add several layers of perceptron.

7. **GoogLeNet Architecture:**
   GoogLeNet, also known as Inception-v1, is a Convolutional Neural Network (CNN) architecture developed by Google. It's renowned for its high accuracy in image classification tasks while keeping computational complexity relatively low. Key features include Inception modules for capturing features at multiple scales efficiently, global average pooling instead of fully connected layers for parameter reduction, 1x1 convolutions for depth increase, and auxiliary classifiers for training stability. Despite its 22-layer depth, GoogLeNet is designed to be easy to train and efficient in terms of computations.

4.2 Transfer-Learning Approach

Transfer learning in deep learning refers to using a pretrained network for a new task. Because it trains the network with less input and great accuracy, transfer learning is highly popular in deep learning. In transfer learning, a computer makes better generalizations about one task by using experience from a prior one. The final few levels of the trained network are replaced with new layers during transfer learning. Examples of these additional layers include a completely connected layer and a SoftMax classification layer, both of which include 38 classes in our paper. We removed the layer's freezing and stacked one activation layer, one batch-normalization layer, and one dropout layer in each model. Different learning rates, batch sizes, and dropout levels.

4.3 Computational Study

In this computational study, we train five state-of-the-art object detection algorithms and eighteen state-of-the-art classification algorithms on the PlantDoc dataset. Before starting the training process, we applied a prepossessing step. At the PlantDoc dataset, there was a total of 28 annotations that needed to be fixed. The bounding boxes were out of frame in some images; thus they were trimmed to match the image's edges.

4.4 Dataset

It seems like you're describing an experiment conducted using the Plant Village dataset for training and testing purposes. Here's a summary of the experiment and the dataset used:

1. **Dataset Description**: The Plant Village dataset consists of 54,305 images of leaves, encompassing both healthy and diseased plants. The dataset includes images from 14 different plant species, categorized into 38 classes based on the presence of diseases. Each class represents a specific combination of plant species and disease.
2. **Dataset Formats**: The experiment utilized three different formats of the Plant Village dataset:
    - Coloured leaf images: Original coloured images of leaves.
    - Segmented leaf photos: Images with segmented leaves and smoothed backgrounds for improved analysis.
    - Grayscale leaf images: Images converted to grayscale format.
3. **Dataset Split**: The dataset was divided into training and testing sets using different splits:
    - 80-20 split: 80% of the data used for training and 20% for testing.
    - 70-30 split: 70% of the data used for training and 30% for testing.
    - 60-40 split: 60% of the data used for training and 40% for testing.
4. **Experimental Setup**: Various batch sizes, learning rates, and dropout values were tested during the experiment. The effectiveness of the adopted strategies was evaluated using the different dataset formats and split ratios.
5. **Evaluation**: Performance evaluation was conducted to assess the effectiveness of the strategies employed. This likely involved metrics such as accuracy, precision, recall, and F1 score, among others, to measure the model's performance on the testing set.
6. **Data- Preprocessing**: Data pre-processing is a fundamental stage in preparing datasets for machine learning. It includes changing raw data into a configuration reasonable for model training. Normal pre-processing procedures incorporate data cleaning to eliminate irregularities or blunders, standardization to scale data inside a particular reach, highlight scaling to guarantee highlights have comparative ranges, and taking care of missing qualities through ascription or evacuation.

Overall, the experiment aimed to explore the impact of different dataset formats and split ratios, as well as hyperparameter settings, on the performance of the trained models for classifying healthy and diseased plant leaves.

Class	Plant Name	Disease Name	Causes Virus Name	Type of Disease	No. of Images
C1	Apple	Healthy	-	-	1645
C2	Apple	Apple scab	*Venturia inaequalis*	Fungus	630
C3	Apple	Black rot	*Botryosphaeria obtusa*	Fungus	621
C4	Apple	Cedar apple rust	*Gymnosp-orangium*	Fungus	275
C5	Blueberry	Healthy	-	-	1502
C6	Cherry	Healthy	-	-	854
C7	Cherry	Powdery mildew	*Podosphaera clandestina*	Biotrophic Fungus	1052
C8	Corn	Healthy	-	-	1162
C9	Corn	Cercospora leaf spot	*Cercospora zeae-maydis*	Fungal	513
C10	Corn	Common rust	*Puccinia sorghi*	Fungus	1192
C11	Corn	Northern Leaf Blight	*Exserohilum turcicum*	Foliar	985
C12	Grape	Healthy	-	-	423
C13	Grape	Black rot	*Guignardia bidwellii*	Fungus	1180
C14	Grape	Esca (Black Measles)	*Phaeomoniella chlamydospora*	Fungus	1383
C15	Grape	Leaf blight (Isariopsis)	*Pseudocercospora vitis*	Fungus	1076
C16	Orange	Healthy	-	-	5507
C17	Peach	Healthy	-	-	360
C18	Peach	Bacterial spot	*Xanthomonas campestris* pv. pruni	Bacterial	2297
C19	Pepper/bell	Healthy	-	-	1478
C20	Pepper/bell	Bacterial spot	*Xanthomonas campestris* pv.	Bacterial	997
C21	Potato	Healthy	-	-	152
C22	Potato	Early blight	*Alternaria solani*	Fungal	1000
C23	Potato	Late blight	*Phytophthora infestans*	Fungal	1000
C24	Raspberry	Healthy	-	-	371
C25	Soybean	Healthy	-	-	5090
C26	Squash	Powdery mildew	*Podosphaera xanthii*	Fungal	1835

Fig 1. Detailed description of plant village datasets with relative information.

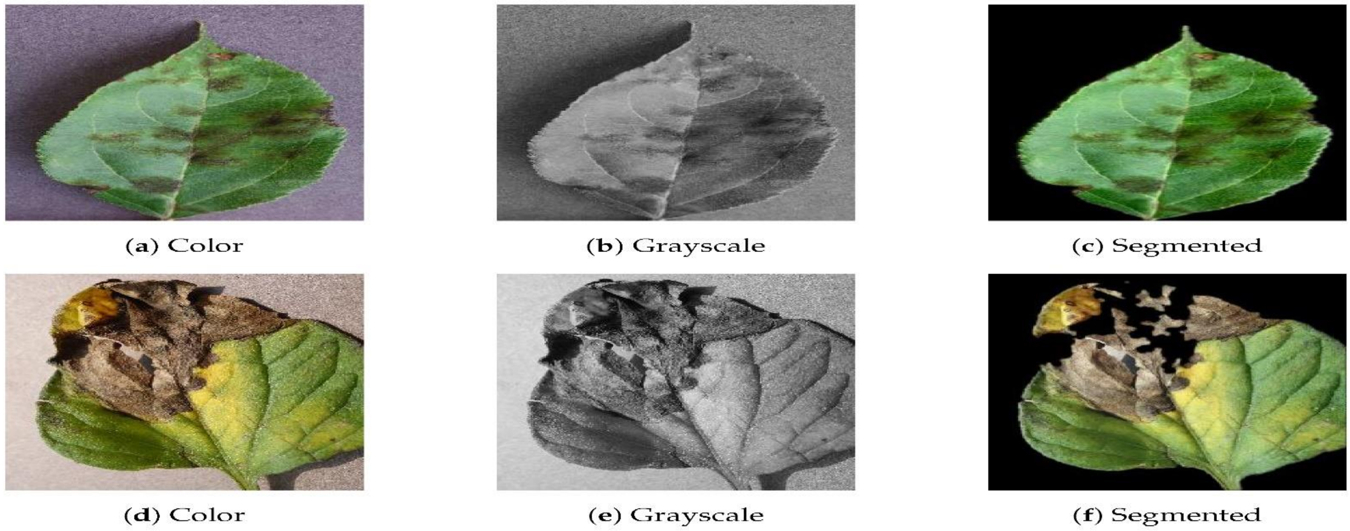

Fig 2. Sample images of colour, grayscale and segmented version of plant village image dataset.

V. RESULT AND DISCUSSION

5.1.RESULT

1] Phase One-Trialling of Image Size

In the first phase, a trial was conducted to evaluate the impact of image size on model performance. In Phase One, image sizes ranging from 155 x 155 to 255 x 255 achieved over 90% accuracy and F1 score. Larger images improved feature extraction but increased processing times. Despite previous recommendations for 224 x 224 resolution, 244 x 244 images yielded the best results.

2] Phase Two-Model Optimization

In Phase Two, the model initially achieved an accuracy of 0.9465 and an F1 score of 0.9359. Before fine-tuning, a plot was analyzed, showing learning rate (on a logarithmic scale) versus loss (refer to Fig. 4). It was observed that there is not a significant decrease in loss between learning rates of 1e-06 and 1e-04. However, a noticeable increase in loss occurs when the learning rate exceeds 1e-04. Considering these findings, several trials were conducted to fine-tune the learning rate.

3] Phase Three-Visualisations

Heat map analysis of the CNN reveals that colour, shape, and texture are essential factors for extracting traits related to plant diseases. Colour, especially, plays a crucial role in distinguishing between similar disorders. The CNN shows strong feature recognition across various species, including rice disease groups.

VI. CONCLUSION AND FUTURE SCOPE

The project aims to develop a comprehensive system with a trained model on a server and a mobile app for identifying diseases in fruits, vegetables, and plants using photos from a phone's camera. This system aims to help farmers quickly identify and treat plant illnesses, making informed decisions about pesticide use. Future research will focus on expanding the model's capabilities to larger geographic areas using drone-captured aerial photographs, potentially leveraging convolutional neural networks for object detection. Drones and autonomous vehicles, including smartphones, will play a crucial role in large-scale disease identification and real-time monitoring in open-field cultivations. Additionally, there is potential for developing an automated pesticide prescription system, which would require approval from an automated disease detection system before farmers can purchase pesticides. This approach could help reduce the excessive and improper usage of pesticides, minimizing their harmful impact on the environment.

VII. REFERENCES

[1] M. E. H. Chowdhury, T. Rahman, A. Khandakar, M. A. Ayari, A.U. Khan, M. S. Khan, N. Al-Emadi, M. B. I. Reaz, M. T. Islam, and S. H. M. Ali, ‘‘Automatic and reliable leaf disease detection using deep learning techniques,’’ AgriEngineering, vol. 3, no. 2, pp. 294–312, May 2021.

[2] M. Tan and Q. Le, ‘‘EfficientNet: Rethinking model scaling for convolutional neural networks,’’ in Proc. Int. Conf. Mach. Learn., 2019, pp. 6105–6114.

[3] A. C. Cruz, A. Luvisi, L. De Bellis, and Y. Ampatzidis, ‘‘X-FIDO: An effective application for detecting olive quick decline syndrome with deep learning and data fusion,’’ Frontiers Plant Sci., vol. 8, p. 1741, Oct. 2017.

[4] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, ‘‘Machine learning in agriculture: A comprehensive updated review,’’ Sensors, vol. 21, no. 11, p. 3758, May 2021.

[5] D.L. Vu , T.K. Nguyen , T.V. Nguyen , T.N. Nguyen , F. Massacci , P.H. Phung , HIT4Mal: hybrid image transformation for malware classification, Trans. Emerging Telecom- mun. Tech no l. 31 (11) (2020) e3789.

[6] Dataset Kaggle, https://www.kaggle.com/thanjaivadivelm.

[7] Chaitali G Dhaware , K H Wanjale , A modern approach for plant leaf disease classification which depends on leaf image processing, in: IEEE International Conference on Computer Communication and Informatics, 2017, pp. 12–16.

[8] P Krithika , S Veni , Leaf disease detection on cucumber leaves using multiclass sup- port vector machine, in: IEEE International Conference on Wireless Communications, Signal Processing and Networking, 2017, pp. 1276–1281.

[9] D. Shah, V. Trivedi, V. Sheth, A. Shah, and U. Chauhan, ‘‘ResTS: Residual deep interpretable architecture for plant disease detection,’’ *Inf. Process. Agricult.*, vol. 9, no. 2, pp. 212–223, Jun. 2022

[10] Robert G. de Luna, Elmer P. Dadios, Argel A. Bandala, “Automated Image CapturingSystem for Deep Learning- based Tomato Plant Leaf Disease Detection and Recognition,”International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD) 2019.