IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Plant Disease Detection Using Deep Learning

[1]Mr.Atharva Santosh Bhagwat, [2]Mr.Tanmay Pramod Garde, [3]Mr.Soham Uday Khedekar,Mr.Rutik Ravindra Vaskar,Prof. Mandar S Joshi
[1]BE Information Technology, [2]BE Information Technology, [3]BE Information Technology, [4]BE Information Technology,[5] Professor
[1]Information Technology,
[1]Finolex Academy Of Management and Technology
, Ratnagiri,India

Abstract: The paper discusses the use of deep learning techniques for the detection and classification of plant diseases from images, highlighting the challenges associated with data collection, preprocessing, and model selection. It provides a comprehensive overview of publicly available datasets for plant disease research and compares the performance of various deep learning models on different datasets. The paper also presents some future directions for research in this field, including transfer learning, multi-spectral imaging, and real-time disease monitoring, emphasizing the potential of deep learning-based approaches for improving plant disease diagnosis and management, which can have significant implications for sustainable agriculture and food security.

***Index Terms* - Agriculture, Plant Diseases, Image Processing, Image Acquisition, Image Pre-Processing, Feature Extraction,Convolutional Neural Network,Random Forest,Ensemble.**

I.INTRODUCTION

PLANT DISEASES CAN SIGNIFICANTLY IMPACT AGRICULTURAL PRODUCTION, LEADING TO ECONOMIC LOSSES AND DECREASED FOOD SECURITY. EARLY DETECTION AND ACCURATE DIAGNOSIS OF THESE DISEASES ARE CRITICAL TO PREVENT THEIR SPREAD AND MINIMIZE THEIR IMPACT. RECENT ADVANCEMENTS IN DEEP LEARNING TECHNIQUES HAVE SHOWN PROMISING RESULTS IN DETECTING AND CLASSIFYING PLANT DISEASES FROM IMAGES. IN THIS PAPER, WE PRESENT A REVIEW OF STATE-OF-THE-ART DEEP LEARNING METHODS USED FOR PLANT DISEASE DETECTION AND CLASSIFICATION. WE DISCUSS THE CHALLENGES ASSOCIATED WITH DATA COLLECTION, PREPROCESSING, AND MODEL SELECTION, AND WE PROVIDE A COMPREHENSIVE OVERVIEW OF PUBLICLY AVAILABLE DATASETS FOR PLANT DISEASE RESEARCH. WE ALSO COMPARE THE PERFORMANCE OF VARIOUS DEEP LEARNING MODELS ON DIFFERENT DATASETS AND HIGHLIGHT THEIR STRENGTHS AND WEAKNESSES. FINALLY, WE PRESENT SOME FUTURE DIRECTIONS AND OPPORTUNITIES FOR RESEARCH IN THIS FIELD, INCLUDING TRANSFER LEARNING, MULTI-SPECTRAL IMAGING, AND REAL-TIME DISEASE MONITORING. OUR STUDY SUGGESTS THAT DEEP LEARNING-BASED APPROACHES HOLD GREAT POTENTIAL FOR IMPROVING PLANT DISEASE DIAGNOSIS AND MANAGEMENT, WHICH CAN HAVE SIGNIFICANT IMPLICATIONS FOR SUSTAINABLE AGRICULTURE AND FOOD SECURITY.

II.PURPOSE

- To detect plant disease from images of leaves of a plant.
- To build a model for disease detection using deep learning.
- To build evaluate different models on the basis of performance to solve the problem.
- To design an Ensemble of models to achieve better detection.

III.LITERATURE SURVEY

After reading and understanding the following research papers we devised the inference methodology to solve this problem which is stated below:

Mohanty et al. (2016): The paper proposes a deep learning-based approach for detecting plant diseases using images. The authors evaluate several CNN models on a large plant disease dataset and compare their performance. They demonstrate that deep learning models can achieve high accuracy rates in detecting and classifying plant diseases.[1]

Fuentes et al. (2018): The paper presents a robust deep learning-based detector for real-time tomato plant disease and pest recognition. The authors use a CNN model with transfer learning and data augmentation techniques to classify images of tomato plants into healthy or diseased/infested categories. The proposed model achieves high accuracy rates in real-time detection of tomato diseases and pests.[2]

Zhang et al. (2017): The paper reviews recent advances in detecting plant diseases using hyperspectral imaging (HSI). The authors discuss the potential of HSI in capturing spectral signatures of plants infected with diseases and present various techniques used for disease detection, such as feature extraction and machine learning algorithms. They highlight the advantages and challenges of HSI for plant disease detection and suggest future research directions.

In summary, these papers demonstrate the potential of deep learning and hyperspectral imaging techniques for detecting and classifying plant diseases. The proposed models achieve high accuracy rates and can be used for real-time disease monitoring, which can have significant implications for sustainable agriculture and food security.

IV. Plant Disease Considerate

To identify the diseases in different plants and learn about features we are considering following plants for the study:

Plant	Plant Disease
Tomato Plant	Tomato Bacterial spot Tomato Early blight Tomato Late blight
Potato Plant	Potato Early blight Late blight
Pepper bell Plant	Pepper bell bacterial spot

Tomato Healthy:

Tomato Bacterial spot:

Tomato Early blight :

Tomato Late blight :

Potato Healthy:

Potato Early blight :

Potato Late blight :


Pepper bell Healthy:


Pepper bell bacterial spot :


V. Methodology

A model is trained to detect the target classes (things to identify in images) using labeled example images. An example of supervised learning is image classification. Raw pixel data was the only input for early computer vision algorithms.

In this project, we will categorize plant leaf photos according to the corresponding disease.

We are employing the Random Forest machine learning method for this.A well-known machine learning method from the supervised learning approach is Random Forest. It can be used for ML issues involving both regression and classification. It is based on the concept of ensemble learning, which is a technique for combining several classifiers to handle challenging problems and improve model performance.

In order to improve the predictive accuracy of the dataset, the Random Forest classifier, as its name suggests, averages several decision trees on various subsets of the input data. The random forest uses predictions from each decision tree and predicts the outcome based on the votes of the majority of projections rather than relying solely on one decision tree.

Higher accuracy and overfitting are prevented by the larger number of trees in the forest.

In order to address this, we are developing a model based on the max voting ensemble approach, which combines the output of various models and provides results as appropriate.

Convolutional Neural Networks (CNNs) are a type of deep learning algorithm that are commonly used in computer vision applications. CNNs are inspired by the visual cortex of animals and are designed to automatically learn and extract features from images, making them highly effective for image recognition, classification, and segmentation tasks.

The key feature of CNNs is their ability to process input data through a series of convolutional layers. Each convolutional layer applies a set of filters to detect local patterns in the input image, which are then passed through non-linear activation functions to introduce non-linearity into the model. Pooling layers are also commonly used to reduce the dimensionality of the output from the convolutional layers. Finally, the output from the convolutional and pooling layers is passed through one or more fully connected layers, which perform the final

classification or regression task. Overall, CNNs have become one of the most widely used deep learning algorithms in computer vision due to their ability to automatically learn and extract relevant features from images, making them highly effective for a wide range of image-related tasks.

The models are built using input data from images of plant leaves with various diseases that have been scaled down to 256x256 and then transformed into R, G, B, and Grey scale to feed to the appropriate models. The aggregator then aggregates these results and provides more accurate results.

The input photos can be resized beyond 256x256, 512x512, etc., which may improve performance, but it is not currently implemented owing to hardware restrictions.

VI. IMAGE PREPROCESSING

Resize and Rescale

Resize - The Application is designed so as to take input images of various devices and thus the resulting sizes of image in pixels may vary . This function resizes the images to a fixed size of 64*64 pixels . Having a smaller size helps in faster execution by 25% percent .

Rescale - This function normalizes the data by converting the values of each pixel from the range of 0 to 255 to 0 to 1 .This makes the model robust to variations and reduces computational costs .

Average Time Taken per Epoch (256*256 pixels) - 13.4 sec

Average Time Taken per Epoch (128*128 pixels) - 12.6 sec

Average Time Taken per Epoch (64 * 64 pixels) - 10 sec

Data Augmentation

The Images can be clicked from different angles and using flips and rotation makes the model robust towards variations in images clicked by the user . Using this improves the performance by 5 percent.

VII. Proposed System

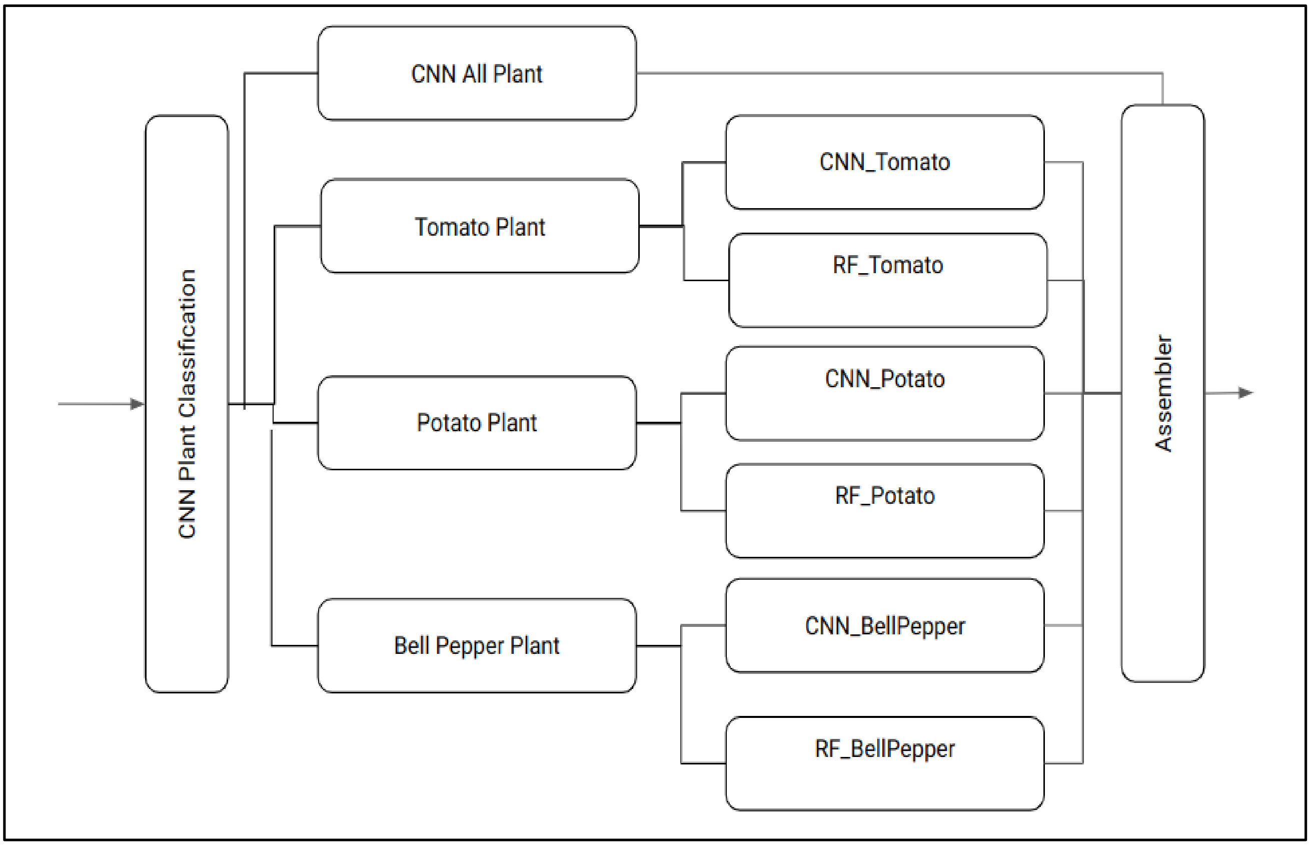
The proposed system is an ensemble of Random Forest models and CNN models for detecting disease in plants using Deep learning techniques.-272414552413

As the above architectural diagram shows there are 5 CNN models and 3 Random Forest models in an Ensemble.The 3 Random Forest Algorithms are working individually for specific plants similarly there are 3 CNN models each one specific to a plant. 1 CNN model is to classify plant image into plant classes. While 1 CNN model is running to classify images of any plant out of them. The Assembler takes input from The Random Forest and CNN from plant class models and the CNN model running on all plants.On basis of maximum voting the disease is detected and output is given by assembler.

This Multilevel ensemble classifies plant type of image at a level 1 at second level the respective models are called based on level 1 result and all plant CNN models.

At level 2 the model calls the one RF model and one CNN model specifically trained to classify that plant.

At the aggregate level the results are calculated using these models output and the disease is predicted with confidence calculated by the aggregator based on results of performance of all the models.



Ensemble Architecture

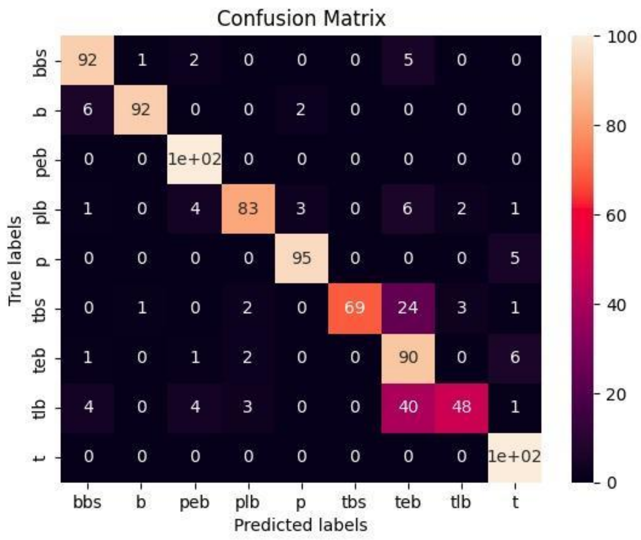
VIII. Results

The Images of Plant were trained on Random Forest Classifiers and CNN models with different combinations of Training Data Sample Size and The Color of the pixel used for Training and Testing . The following are the Results of this testing -

1)ALL_PLANT_CNN_MODEL:

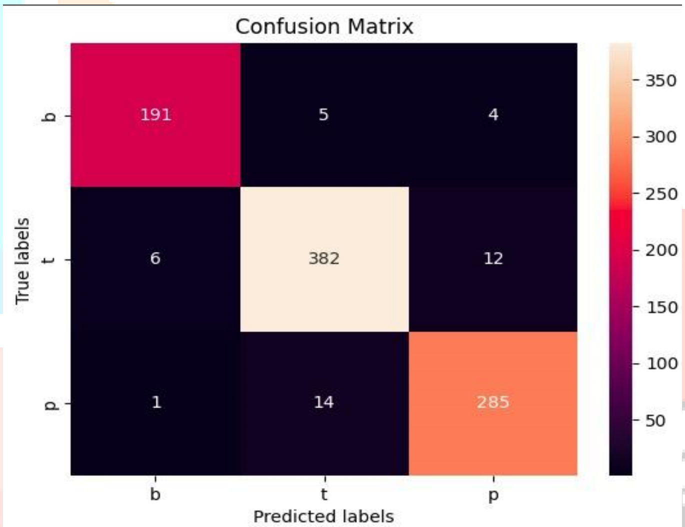
This model detect the plant disease from 9 classes namely (Tomato Bacterial spot,Tomato Early blight ,Tomato Late blight,Potato Early blight,Late blight ,Pepper bell bacterial spot,Tomato healthy,Potato Healthy,Pepper bell Healthy) which on testing on testing dataset its performance metrics are:

Accuracy - 85.44%



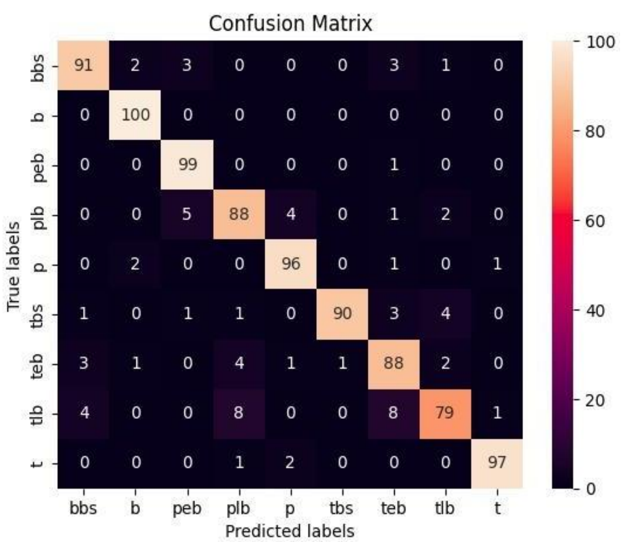
2)Plants Classifier (level1):

This model identifies the leaf to which type of plant (Tomato,Potato,Pepper Bell)it is which following performance metrics:
Accuracy - 95.33%



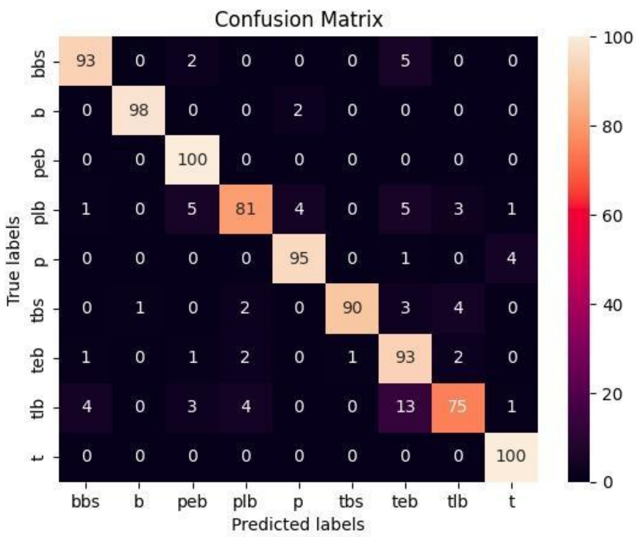
3)Separate classifier (level 2):

The combined model of all plants having one RF and one CNN per plant identifies the disease to a particular plant.
It's Performance Metrics are:
Accuracy- 92.0%
Combined Accuracy- (Hierarchy model) 87.70%



4)Multilevel Ensemble Model:

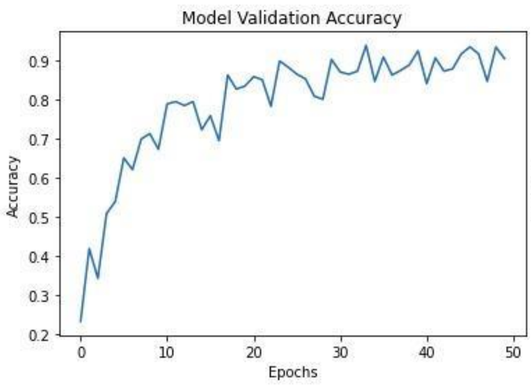
All above models combined in a multilevel/hierarchical models performance metrics are:
Accuracy - 91.66%

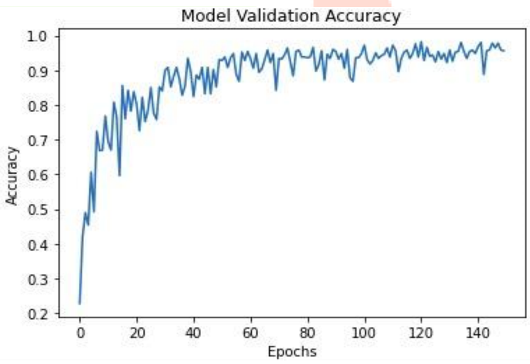


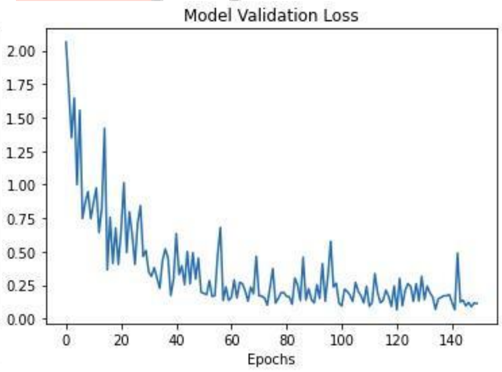
* On the basis of experimentation performed during this research following observations are made.
*Accuray may have variations of up to 2% due to random sampling used in Random forest classifier.

IX. Inference

The image classification system uses two approaches where the hierarchical classifier approach performs better than the direct classifier . The direct model is unable to distinguish diseases of Tomato plants hence when both models classify the plant as “Tomato” using a separate model for that plant helps improve the performance. However, doing an ensemble improves the classification accuracy . The CNN model performance stagnates after 60 epochs .Thus using a model with 50 epochs gives faster performance by 34% and an accuracy loss of 1.7% .







References

[1] Using Deep Learning for Image-Based PlantDisease Detection Sarada Prasanna Mohanty 1,2, David Hughes3,4,5, and Marcel Salathé1,2,6

[2] Fuentes AF, Yoon S, Lee J, Park DS. High-Performance Deep Neural Network-Based Tomato Plant Diseases and Pests Diagnosis System With Refinement Filter Bank. Front Plant Sci. 2018 Aug 29;9:1162. doi: 10.3389/fpls.2018.01162. PMID: 30210509; PMCID: PMC6124392.

[3] "Experimentation with Plant Disease Detection Using ML", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.9, Issue 11, page no. ppa954-a960, November-2022, Available at : http://www.jetir.org/papers/JETIR2211108.pdf