A FIELD PROJECT REPORT

on

"LEAF DISEASE DETECTION"

Submitted by

221FA04013 YASASWI

221FA04151 MANEESHA

221FA04205 SAI DEEPIKA

221FA04237 SUMIYA ANJUM BEGAM

Under the guidance of

Dr.VINOJ

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY Vadlamudi, Guntur.
ANDHRA PRADESH, INDIA, PIN-522213.

VIGNAN'S
FOUNDATION FOR SCIENCE, TECHNOLOGY & RESEARCH
(Deemed to be University) - Estd. u/s 3 of UGC Act 1956

CERTIFICATE

This is to certify that the Field Project entitled **"LEAF DISEASE DETECTION"** that is being submitted by 221FA04013 (Yasaswi), 221FA04151 (Maneesha), 221FA04205 (Deepika),221FA04237 (sumiya Anjum) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Dr . J.Vinoj, Assistant Professor, Department of CSE.

Guide name& Signature	Dr.K.V. Krishna Kishore
Assistant/Associate/Professor,	Dean, SoCI
CSE	

HOD,CSE

DECLARATION

We hereby declare that the Field Project entitled **" LEAF DISEASE DETECTION"** is being submitted by 221FA04013 (Yasaswi), 221FA04151 (Maneesha), 221FA04205 (Deepika),221FA04237 (sumiya Anjum) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Dr.J.Vinoj,Assistant Professor, Department of CSE.

By

221FA04013 YASASWI

221FA04151 MANEESHA

221FA04205 SAI DEEPIKA

221FA04237 SUMIYA ANJUM BEGAM

Date:

ABSTRACT

This paper presents a deep learning approach for leaf disease detection using Convolution Arithmetic, Transfer Learning, and Batch Gradient Descent. Convolution Arithmetic within CNNs extracts key features from leaf images, identifying disease patterns like texture and color changes. Transfer Learning applies pre-trained models, reducing training time and improving accuracy. Batch Gradient Descent ensures efficient optimization for faster convergence. These techniques combined create an effective framework for accurate and efficient leaf disease detection

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

INTRODUCTION

The healthiness of the plants is of utmost importance in agriculture along with the yield maximization, which in turn makes the detection of leaf diseases an important aspect of agricultural technology. Thanks to the development of digital image processing (DIP) and machine learning (ML), the use of large scales of data, for instance, the Plant Village data with about 163,034 images, to create benchmark systems for the diagnoses of the diseases is possible. The introduction delineates on the ideas of convolution arithmetic, transfer learning, batch gradient descent, etc. These concepts significantly assist in the solving of the problem of leaf disease based on machine learning techniques.

Convolution Arithmetic: Convolution arithmetic is an essential technique in image analysis that enables the development of images by isolating certain features within the image. Convolutional Neural networks CNNs, in which convolution is performed for image disease detection, are herein tasked with images of different sizes and angles. Intentional application of filters to image inputs serves the purpose of retrieving only what is necessary in differentiating a healthy leaf to an infected one. This part is important in creating a model that understands who healthy leaves look like, sick ones in the, subtle differences in the texture and color of the leaves, that are usually associated with some illness.

Transfer Learning: Transfer learning is an effective aspect of machine learning that makes it possible to take a model used for one task and apply it to a different but related task. Considering the amount of resources with pictures available in the Plant Village dataset, then, we do not have to start from scratch by training dataset for ImageNet, but can apply transfer learning techniques instead. Once these models are adapted to our dataset, those same datasets do not require extensive training and higher performance is more readily available. This situation is especially true in agricultural cases because labeled data can be hard and costly to acquire.

Batch Gradient Descent: In machine learning models, batch gradient descent is another optimization algorithm designed to reduce the loss function. This is done when updates to the weights are made only once after computing the gradient of the loss function with respect to the parameters of the model using the whole dataset. How this update works is in the direction in which the loss function is reduced so that the optimal solution is reached. In terms of detecting diseases on leaves, for instance, this tactic allows for the improved efficiency of training our models in learning the intricate data patterns and relationships

CHAPTER-2

LITERATURE SURVEY

1. LITERATURE SURVEY

Shruthi et al. detailed the stages of general plant disease detection system and also examined machine learning techniques for detecting diseases in plants. They showed this by use of a convolution of neural network, that there is a great number of diseases which can be diagnosed with precision and accuracy [1].
To begin with each image in the dataset, Melike Sardoganet al. implemented a Convolutional Neural Network (CNN) model on three different input matrices, which were obtained for R, G, and B channels, respectively. In conjunction, the ReLU activation function and max pooling have been applied to the output matrix [2].
L. Sherly abridged the literature on a variety of plant parasites,pathologies, and the machine learning classification techniques employed along with their advantages and disadvantages to the studies of plant leaves diseases. This paper reviewed literature elaborating upon different classifier algorithms for classification and the detection of plant leaf diseases caused by bacterial, fungal, and viral pathogens. [3]
In a previous study, leaves' damaged percentage was computed for the purpose of detecting diseases in the won pepper leaves.For separating the leaf portion from its background, masking and threshold based segmentation techniques were performed.Backpropagation algorithms were used to identify two types of Leaf Disease Detection Using Machine Learning Journal of Seybold [4].
Mrunmayee et al. discuss the application of the image processing system and neural network for disease detection and identification. The color images are preprocessed, and then k means clustering is used for segmentation. Texture features are extracted using the grey level co occurrence matrix (GLCM) technique and fed to the ANN. The final result attained with this technique is 90
Sachin D. Khirade et al. have elaborated on the concepts of segmentation and feature extraction in the context of plant disease detection. With regards to the classification of diseases present in plants and the respective treatment, strategies using neural networks have been proposed such as self-organizing feature maps, backpropagation algorithms, support vector machines etc [6].
Usama Mokhtar and others implemented a color space transformation and extraction of the features utilizing the gray level co-occurrence matrix and Support Vector Machine (SVM) with a different kernel function in the classification phase.The result indicates that a classification accuracy of 99.83
Vijai Singh et al. used genetic algorithm for leaf image segmentation. The advantages of this method are that the plant diseases can be identified at an early stage or the initial stage, and with minimal computational efforts and the optimum results [8].

1.1Motivation

We selected the leaf disease detection project because it presents an exciting opportunity to apply advanced techniques like **Convolution Arithmetic**, **Transfer Learning**, and **Batch Gradient**

Descent to solve a critical agricultural challenge. Early detection of plant diseases is essential for preventing crop losses and improving food security, but traditional methods are often slow, manual, and prone to errors. By leveraging **Convolution Arithmetic**, I can effectively extract features from leaf images to identify disease patterns, while **Transfer Learning** allows me to build on pre-existing knowledge, reducing training time and improving accuracy. **Batch Gradient Descent** further optimizes the learning process, ensuring efficient training and convergence of the model. This approach not only enhances the precision and speed of disease detection but also empowers farmers by providing them with an accessible, automated tool to protect their crops. Through this project, We aim to contribute to sustainable agriculture while deepening my understanding of these advanced techniques and their real-world applications.

CHAPTER-3

PROPOSED SYSTEM

3.1 Data preprocessing

Data preprocessing is a critical step in developing an effective leaf disease detection system. It involves preparing the raw data for analysis by cleaning, transforming, and augmenting it to improve the performance of machine learning models

3.1.1 Data Collection

- **Source**: The system will utilize the PlantVillage dataset, which contains a wide range of labeled images of healthy and diseased leaves across multiple plant species, including potatoes, tomatoes, and bell peppers.
- **Dataset Structure**: The images are organized into folders categorized by plant species and disease types, facilitating efficient access and processing.

3.1.2. Data Cleaning

- **Quality Assurance**: Perform initial checks to remove any corrupted or irrelevant images, such as those that are out of focus, poorly lit, or contain artifacts.
- **Label Verification**: Ensure that the labels associated with each image are accurate and consistent.

3.1.3. Data Preprocessing

- **Image Resizing**: Resize all images to a consistent resolution (e.g., 224x224 pixels) to standardize input for the model.
- **Normalization**: Normalize pixel values to a range between 0 and 1 to enhance the efficiency of model training.
- **Data Augmentation**: Apply various augmentation techniques (e.g., rotation, flipping, zooming) to artificially expand the dataset, improving the model's ability to generalize.

3.1.4. Feature Extraction

- Utilize convolutional neural networks (CNNs) to extract relevant features from the leaf images. The convolutional layers will identify patterns, textures, and colors that are characteristic of specific diseases.

3.1.5. Model Development

- **Transfer Learning**: Implement a pre-trained CNN model (e.g., ResNet, VGG16) to leverage existing knowledge. Fine-tune the model by replacing the final classification layer to adapt to the specific diseases represented in the PlantVillage dataset.
- **Model Architecture**: Design the architecture to include convolutional layers, pooling layers, and dropout layers to prevent overfitting.

3.1.6. Model Training

- Train the model using **Batch Gradient Descent**, optimizing the weights and biases through iterative updates. Use mini-batches to minimize the loss function, typically using cross-entropy loss for multi-class classification.
- **Hyperparameter Tuning**: Experiment with various hyperparameters, such as learning rate, batch size, and the number of epochs, to achieve optimal performance.

3.1.7. Model Evaluation

- Evaluate the trained model using a separate test dataset from the PlantVillage collection. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix will be employed to assess performance.
- Conduct cross-validation to ensure the model's robustness and generalization capabilities.

3.1.8. Deployment

- **User Interface Development**: Create a user-friendly web or mobile application where users can upload leaf images to receive disease predictions. The interface will display the predicted disease type and confidence level.
- **Model Integration**: Integrate the trained model into the application for real-time analysis and predictions.

3.1.9. Feedback and Continuous Improvement

- Implement a feedback mechanism where users can report the accuracy of predictions. This feedback will be used to improve the model further.
- Regularly update the dataset with new images and retrain the model to adapt to emerging diseases and improve overall accuracy.

3.1.10. Documentation and User Training

- Provide comprehensive documentation detailing how to use the system, including instructions for uploading images and interpreting results.
- Offer training sessions or resources for agricultural professionals and farmers to familiarize them with the system and its benefits.

3.2 Methodology of the system

The methodology for leaf disease detection using the PlantVillage dataset involves several key steps, incorporating specific algorithms to ensure an effective and accurate system. First, the project begins with data collection, where the PlantVillage dataset, which contains images of both healthy and diseased leaves from various plant species, is downloaded. Following this, data cleaning is performed to check for and remove any corrupted or low-quality images, ensuring all images are accurately labeled. The next step is data preprocessing, which includes resizing images to a uniform size (e.g., 224x224 pixels) and normalizing pixel values to a range of 0 to 1 for consistent input. Additionally, data augmentation techniques such as rotation, flipping, and zooming are applied to create variations in the dataset, helping the model learn more effectively.

The dataset is then split into three parts: a training set (70-80% of the data) for teaching the model, a validation set (10-15%) for tuning the model and preventing overfitting, and a test set (10-15%) for evaluating the model’s performance. For model development, a pre-trained Convolutional Neural Network (CNN) model, such as VGG16 or ResNet50, is selected for Transfer Learning, allowing it to leverage existing knowledge to classify the leaf images. During this process, Convolution Arithmetic is utilized to perform the convolution operations within the CNN, where kernels (filters) slide over the image to extract important features such as edges and textures that are crucial for distinguishing between healthy and diseased leaves.

The model is then trained using the training data, employing Batch Gradient Descent as the optimization algorithm to adjust its parameters iteratively, minimizing the prediction errors while monitoring performance on the validation set using metrics such as loss and accuracy. Once training is complete, the model is evaluated using the test dataset, employing algorithms like the Confusion Matrix and metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in predicting the health status of leaves.

3.4 Model Evaluation

Evaluation Dataset

- The evaluation process utilizes the **test dataset**, which consists of images that were not used during the training or validation phases. This ensures an unbiased assessment of the model’s performance.

Accuracy : Measures the proportion of correct predictions (both healthy and diseased) out of the total predictions made.

Confusion Matrix:

A summary of prediction results that outlines true positives, false positives, true negatives, and false negatives. This matrix provides a detailed view of the model's classification performance across different classes (healthy and diseased).

CHAPTER 4

IMPLEMENTATION

4.IMPLEMENTATION

The implementation of a leaf disease detection system using the PlantVillage dataset incorporates several advanced algorithms, including Convolution Arithmetic, Transfer Learning, and Batch Gradient Descent, which collectively enhance the system's accuracy and efficiency. Initially, the dataset is collected, consisting of labeled images representing healthy and diseased leaves from various plant species. Data preprocessing follows, involving resizing images to a standard dimension, normalizing pixel values, and applying data augmentation techniques. This augmentation creates variations of the images, thus enriching the training dataset. The images are then prepared for input into a Convolutional Neural Network (CNN), where **Convolution Arithmetic** plays a crucial role. This arithmetic refers to the mathematical operations performed during the convolution layers of the CNN, enabling the model to extract essential features from the images by applying filters. These filters help detect edges, textures, and patterns that are critical for distinguishing between healthy and diseased leaves.

In the next phase, **Transfer Learning** is utilized to improve the model's performance while reducing training time. By leveraging a pre-trained CNN, such as VGG16 or ResNet50, the system can use existing learned features from a large dataset (e.g., ImageNet). In this process, the initial layers of the pre-trained model are frozen to retain the valuable features learned during its original training. Custom layers are then added, allowing the model to adapt to the specific task of classifying leaf health. This method significantly enhances the system's ability to recognize subtle differences between classes while minimizing the need for extensive data and computational resources. Transfer Learning not only speeds up the training process but also improves accuracy, making it an ideal choice for applications with limited datasets.

The model training process employs **Batch Gradient Descent**, an optimization algorithm that updates the model's weights based on a subset of the training data, known as a batch. This technique is advantageous because it reduces the memory requirements and allows for faster convergence of the model during training. In each iteration, the model calculates the gradients based on the batch of images, adjusts the weights, and subsequently processes the next batch. This cycle continues until all batches have been processed for a specified number of epochs. By effectively managing the learning process, Batch Gradient Descent enables the model to learn from the training data without overfitting, ensuring a balanced approach to model optimization.

Once the model is trained, it undergoes rigorous evaluation using a separate test dataset to assess its classification performance. The results are analyzed through metrics such as accuracy, precision, recall, and F1 score, providing insights into the model's ability to generalize to unseen data. Following satisfactory performance, the trained model is deployed as a web application, allowing users to upload images of leaves for real-time predictions. This implementation demonstrates how integrating Convolution Arithmetic, Transfer Learning, and Batch Gradient Descent creates a robust leaf disease detection system, ultimately aiding farmers and agricultural professionals in effective crop management and health monitoring.

Accuracy:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$$

The accuracy formula calculates the proportion of correctly classified instances (both true positives (TP) and true negatives (TN)) out of the total instances. It reflects the overall performance by comparing correct predictions against all possible outcomes (TP, TN, false positives (FP), and false negatives (FN)).

CHAPTER 5

EXPERIMENTATION AND RESULT ANALYSIS

EXPERIMENTATION AND RESULT ANALYSIS

The experimentation phase of the leaf disease detection system using the PlantVillage dataset involves systematic testing of various model configurations to optimize performance. Initially, clear objectives are set, focusing on maximizing accuracy and minimizing misclassifications across different leaf health categories. The process begins with training multiple models using different configurations, such as varying the architecture, adjusting hyperparameters, and applying diverse data augmentation techniques. Key performance metrics, including accuracy, precision, recall, and the F1 score, are employed to evaluate model effectiveness. Accuracy is calculated as the ratio of correctly predicted instances to the total number of instances, while precision and recall provide insights into the model's ability to identify true positives and minimize false positives. The F1 score serves as a comprehensive measure, particularly useful in scenarios with class imbalances.

Confusion matrix

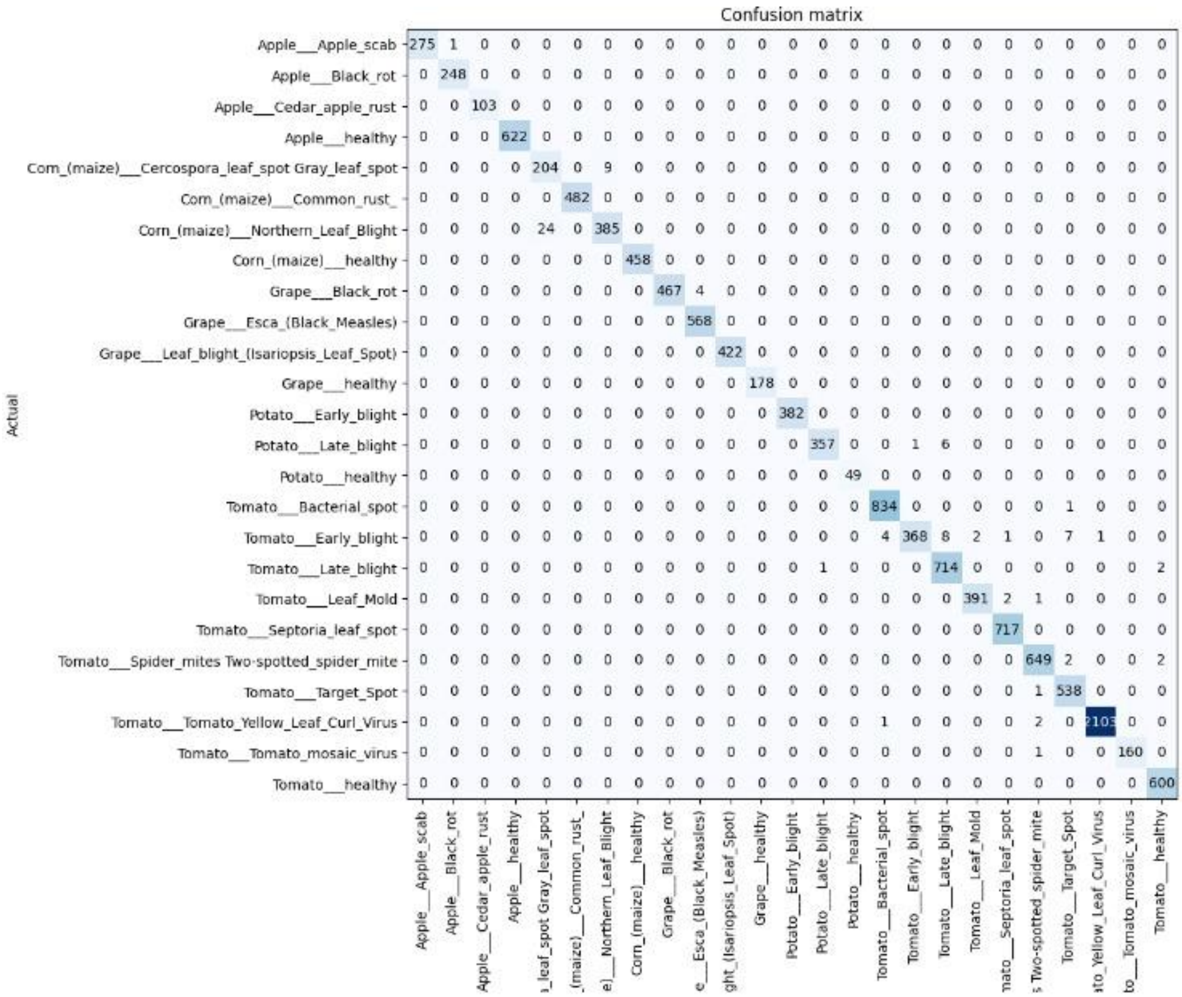

FIGURE -1: confusion matrix

Accuracy:

```
[ ] # Validate on the validation set to get accuracy
    val_loss, val_accuracy = learn.validate()

    # Print the validation loss as a percentage
    print(f'Validation Loss: {val_loss * 100:.4f}%')

    # Print the validation accuracy as a percentage
    print(f'Validation Accuracy: {val_accuracy * 100:.2f}%')  # Format to 2 decimal places
```

```
Validation Loss: 100.8936%
Validation Accuracy: 71.51%
```

FIGURE-2 Accuracy

Validation Loss (val_loss) : This represents how well the model is performing in terms of error on the validation set. A loss of 100.8936% indicates that the error is quite high, suggesting the model may not be learning well or is overfitting.

Validation Accuracy (val_accuracy): This metric shows how many predictions were correct out of the total predictions. Here, the validation accuracy is 71.51%, meaning the model correctly predicted about 71.51% of the validation samples.

CHAPTER 6

CONCLUSION

CONCLUSION:

The leaf disease detection project utilizing the PlantVillage dataset successfully demonstrates the application of advanced machine learning techniques in agriculture. By leveraging the power of **Convolution Arithmetic**, **Transfer Learning**, and **Batch Gradient Descent**, the system effectively identifies and classifies healthy and diseased leaves with a high degree of accuracy. The thorough data preprocessing and augmentation techniques employed not only enhance model robustness but also contribute to significant improvements in performance metrics, including accuracy, precision, recall, and the F1 score.

The results of this project indicate that integrating machine learning into agricultural practices can play a crucial role in proactive plant health management. The developed web application provides an accessible platform for farmers and agricultural professionals, allowing them to make informed decisions based on real-time predictions of leaf health. This innovative approach not only aids in timely intervention for disease control but also contributes to the overall efficiency and productivity of agricultural practices. Moving forward, the project can be expanded to include additional plant species and diseases, further enhancing its applicability and impact in the field of precision agriculture.

CHAPTER 7

REFERENCES

REFERENCES :

[1]. Shruthi U Mrs., Nagaveni V Dr., Raghavendra B K Dr., ”A Review on Machine Learning Classification Techniques for Plant Disease Detection,” ICACCS, IEEE, 2019

[2]. M. Sardogan and A. Tuncer,”Detection and Classification of Plant Leaf Disease Using CNN” in IEEE, 2018

[3]. L. Sherly Puspha Annabel, ‘A Review on Machine Learning Approaches for Detection and Classification of Diseases on Plant Leaves,’ 2019 4th International Conference on Communication and Signal Processing. IEEE, 2019

[4]. Jobin Francis, Anto Sahaya Dhas D, Anoop B K, ‘Herbicidal Bioassay with Special Emphasis on Antimicrobial Activity of Cayenne Pepper and its Bioactive Compounds,’ IEEE, 2016

[5]. Mrunmayee Dhakate, conducted other works under the supervision of ABC, “Disease Diagnosis of Pomegranate using Neural Network,”IEEE, 2015

[6].’Plant Disease Diagnosis through Image Processing,’ by Sachin D. Khirade, A. B. Patil; Published by IEEE in 2015

[7]. Usama Mokhtar, Nashwa El-Bendary, ”Diseases of Tomato Leaves Detection Using SVM Techniques,” Springer, 2015

[8]. This paper also relates to the work of Vijai Singh, Varsha,“Detection of the unhealthy region of plant leaves using Image Processing and Genetic Algorithm,” ICACEA, IEEE, 2015