A FIELD PROJECT REPORT

on

Potato Plant Leaf Disease Detection Using

Hybrid Transformer with ResNet50

## ABSTRACT

Agriculture stands as a cornerstone of livelihoods worldwide, providing sustenance, economic stability, and ecological balance. Yet, the vulnerability of crops to diseases poses significant threats to agricultural productivity. Plant disease identification is essential since a single leaf's health can have a ripple effect on the crop as a whole, affecting yield and quality. Our research addresses this critical issue by proposing a novel solution leveraging deep learning models. In particular, we integrate ResNet50, a convolutional neural network renowned for its ability to recognize and classify images, with a Vision Transformer, a transformer architecture adept at handling image data. Through experimentation, we achieved promising results: ResNet50 yielded an accuracy of 98.3%, while Vision Transformer achieved 51.3% accuracy. Notably, our optimized solution, combining ResNet50 features with the Vision Transformer, achieved an impressive accuracy of 99.3%. By detecting plant diseases early and accurately, our approach offers farmers valuable insights into disease severity, enabling proactive measures to safeguard crop health and enhance agricultural sustainability.

**Index terms** — Agriculture, Plant disease detection, Deep learning, ResNet50, Vision Transformer, Convolutional neural network, Transformer architecture, Image classification, Crop health, Agricultural sustainability.

**TABLE OF CONTENTS**

1. [Introduction 1](#_bookmark0)
   1. [What is traffic congestion and what causes it? 2](#_bookmark1)
   2. [The consequences of traffic congestion 3](#_bookmark2)
   3. [The economic and environmental effects of traffic congestion 3](#_bookmark3)
   4. [Current Methodologies 4](#_bookmark4)
   5. Applications of ML to combat congestion 5
2. [Literature Survey 7](#_bookmark5)
   1. [Literature review 8](#_bookmark6)
   2. [Motivation 10](#_bookmark7)
3. [Proposed System 12](#_bookmark8)
   1. [Input dataset 14](#_bookmark9)
      1. Detailed features of dataset 14
   2. [Data Pre-processing 15](#_bookmark10)
      1. [Missing values 16](#_bookmark11)
         1. [Parameters of the fillna method 17](#_bookmark12)
      2. [Data Encoding 17](#_bookmark13)
   3. [Model Building 18](#_bookmark14)
      1. [MLP Algorithm 19](#_bookmark15)
   4. [Methodology of the system 21](#_bookmark16)
   5. [Model Evaluation 24](#_bookmark17)
   6. [Constraints 26](#_bookmark18)
   7. Cost and Sustainability Impact 27
4. [Implementation 29](#_bookmark19)
   1. Environment Setup 30
   2. Sample code for preprocessing and MLP operations 31
5. Experimentation and Result Analysis 33
6. Conclusion 41
7. References 43

**LIST OF FIGURES**

|  |  |
| --- | --- |
| Figure 1. Architecture of the proposed system | 13 |
| Figure 2. Various features in the dataset | 14 |
| Figure 3. Pre-processing Overview | 15 |
| Figure 4. Missing Values | 16 |
| Figure 5. Architecture of ML | 18 |
| Figure 6. Traffic volume across various weather condition | 34 |
| Figure 7.. Traffic Volume by Time of Day | 33 |
| F igure 8. Correlation Matrix | 35 |
| Figure 9. Comparison of R2 Scores for Different Models | 36 |
| Figure 10. Comparison of MAPE Values for Different Models and Intervals | 37 |
| Figure 11. Effect of various hyper parameters over MAPE and R2 Scores | 38 |

**LIST OF TABLES**

Table 1.Comparision results 14

# CHAPTER-1 INTRODUCTION

### INTRODUCTION

Worldwide, agriculture is the backbone of livelihoods because it sustains people, maintains the economy, and maintains ecological equilibrium. Across diverse landscapes and cultures, agriculture plays a fundamental role in meeting the nutritional needs of populations and driving socio-economic development. From smallholder farms to large-scale commercial operations, the cultivation of crops sustains livelihoods, fosters community resilience, and supports global food security. The very foundation of agricultural prosperity faces formidable challenges, chief among them being the persistent threat of plant diseases.

In today's agricultural landscape, farmers grapple with an array of challenges posed by plant leaf diseases, which significantly impact crop health, productivity, and economic viability. The emergence and spread of diseases such as blights, wilts, and rot can decimate entire harvests, leading to devastating consequences for farmers and communities alike. Beyond the immediate loss of yield, plant diseases also incur substantial economic costs through reduced marketability of produce, increased input expenditures on disease management, and long-term degradation of soil health. Moreover, the repercussions extend beyond individual farms to ripple throughout entire agricultural systems, exacerbating food insecurity and undermining rural livelihoods.

Amidst these challenges, the integration of deep learning techniques offers a promising avenue for combating plant diseases and safeguarding crop health. By harnessing the power of artificial intelligence, researchers and practitioners can leverage vast datasets comprising images of diseased plants to develop predictive models capable of accurately identifying and classifying diseases. Convolutional neural networks (CNNs), such as ResNet50, have emerged as indispensable tools for image recognition and classification, enabling the automated analysis of complex visual data.

Our proposed model represents a novel approach to plant disease detection, leveraging the complementary strengths of ResNet50 and the Vision Transformer. By combining the robust image recognition capabilities of ResNet50 with the contextual understanding afforded by the Vision Transformer, we aim to enhance the accuracy and efficiency of disease detection in potato plants. Specifically, our model integrates ResNet50 to extract informative features from leaf images, which are subsequently fed into the Vision Transformer for further analysis. Through this synergistic approach, we seek to achieve superior performance in disease detection, thereby empowering farmers with timely insights into disease severity and enabling proactive measures to safeguard crop health and enhance agricultural sustainability.

# CHAPTER-2 LITERATURE SURVEY

## LITERATURE SURVEY

#### Literature review

Hasibuk islam peyal et al.[1] proposed a lightweight 2D CNN approach for classifying 14 classes of tomato and cotton crop diseases. Their Android application, "Plant Disease Classifier," achieved an average accuracy of 97.36%, outperforming several pre-trained models. Despite its lightweight nature, the model demonstrated superior performance compared to VGG16, VGG19, InceptionV3, MobileNet, and MobileNetV2.

Wassawa Shafik et al. [2] conducted a systematic literature review on plant disease detection methods, focusing on vision-based AI, ML, and DL approaches. SVMs and LR classifiers showed improved accuracy, but disease localization remains challenging. Cognitive CNNs with attention mechanisms are emerging trends. Limited availability of large datasets and models suitable for small devices poses a limitation, urging the need for robust solutions accommodating various crops and diseases.

R.Satya Rajendra Singh et al.[3] proposed a novel framework, 'Zero-Shot Transfer Learning,' addressing the challenge of classifier performance when trained on a source domain and tested on a target domain with different data distributions. This framework, exemplified by tomato and potato datasets, incorporates CNN models, data augmentation, synthetic data generation, and discriminative losses, enhancing classifier performance in zero-shot scenarios.

Khalid M.Hosny et al.[4] proposed a novel lightweight deep convolutional neural network (CNN) model augmented with traditional handcrafted local binary pattern (LBP) features for plant leaf disease classification. The model was trained and tested on publicly available datasets (Apple Leaf, Tomato Leaf, and Grape Leaf), achieving validation accuracies of 99%, 96.6%, and 98.5%, and test accuracies of 98.8%, 96.5%, and 98.3%, respectively.

Md.Sakib Hossain Shovon et al.[5] proposed "PlantDet," a deep ensemble model incorporating InceptionResNetV2, EfficientNetV2L, and Xception architectures, addressing underfitting and overfitting issues for rice and betel leaf disease classification. PlantDet achieved state-of-the-art performance, surpassing previous models with 98.53% accuracy for rice leaf diseases and outperforming baseline models for betel leaf disease classification. However, limitations may include scalability concerns with larger datasets and potential challenges in generalizing to diverse environmental conditions.

Xin Zhang et al. [6] proposed method integrates capsule networks and enhanced residual networks (ResNet). They optimize ResNet by replacing its initial convolutional layer with a concatenation of smaller kernels and attention mechanism within residual blocks to enhance feature extraction. This improved ResNet is then integrated with CapsNet, with modifications to preserve positional information. Evaluation on multiple datasets showcases the proposed SE-SK-CapResNet model achieving remarkable accuracy rates of 98.58%, 95.08%, and 97.19% across different datasets.

Diana Susan Joseph et al. [7] addressed and developed datasets for rice, wheat, and maize, targeting common diseases affecting crop yields. Applying eight fine-tuned deep learning models, they achieved notable testing accuracies, with Xception and MobileNet excelling in maize leaf disease recognition, achieving testing accuracies of 0.9580 and 0.9464, respectively.They proposed a novel CNN model trained from scratch, exhibiting promising results across all three food grain datasets with testing accuracies of 0.9704, 0.9706, and 0.9808 for maize, rice, and wheat, respectively.

Vasileios et al. [8] introduced extensive computational analysis on the dataset, they found YOLOv5 to exhibit high accuracy in object detection. For image classification, ResNet50 and MobileNetv2 demonstrated the most optimal trade-off between accuracy and training time, with ResNet50 achieving a total accuracy of 61.01% and MobileNet achieving a total accuracy of 59.74%.

Serosh Karim Noon et al. [9] introduced a modified Spatial Pyramid Pooling (SPP) layer to enhance feature extraction across various scales and employ skip connections to improve generalization capability. They utilized an IoU-based regression loss function.Experimental results on a dataset with the proposed model achieving a mean average precision of 73.13% and 327% better test accuracy compared to the original YOLOX model.

Muhammad Hammad Saleem et al. [10] propose an optimized version of the Region Based Concolution Neural Network(RFCN) deep learning model and evaluate various techniques for improved performance. Through empirical observations and cross-validation, they achieve a mean average precision of 93.80%, surpassing default settings by 19.33%.

Chunduri Madhurya et al. [11] proposed an optimized framework based on YOLOv7, termed YR2S, is developed to address limitations in plant leaf disease detection.Utilizing pre-processing and hybrid optimization techniques, including PCFAN for feature map generation and ShuffleNet with ERSO for classification optimization, the framework achieves high accuracy (99.69%).

#### 

# CHAPTER-3 PROPOSED SYSTEM

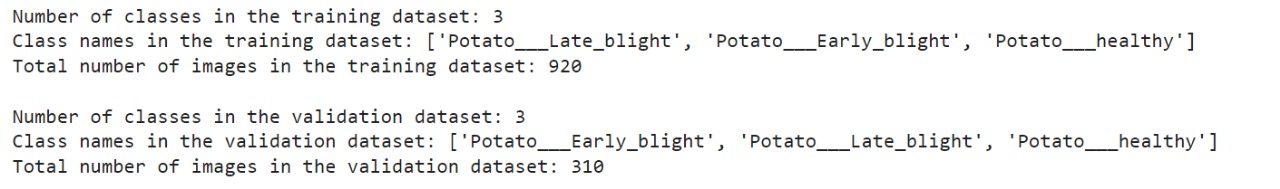
### PROPOSED SYSTEM

A diagram of a software process

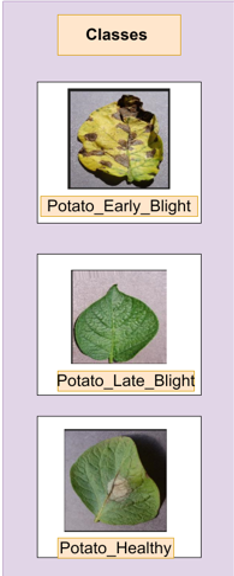
Description automatically generated

#### Input dataset

In the Dataset, we have a collection of images depicting potato plant leaves, each labeled with one of three categories: Potato\_\_\_Early\_blight, Potato\_\_\_healthy, and Potato\_\_\_Late\_blight. These labels correspond to different stages of plant health, ranging from early blight symptoms to late blight infestations, as well as images representing healthy potato plants.



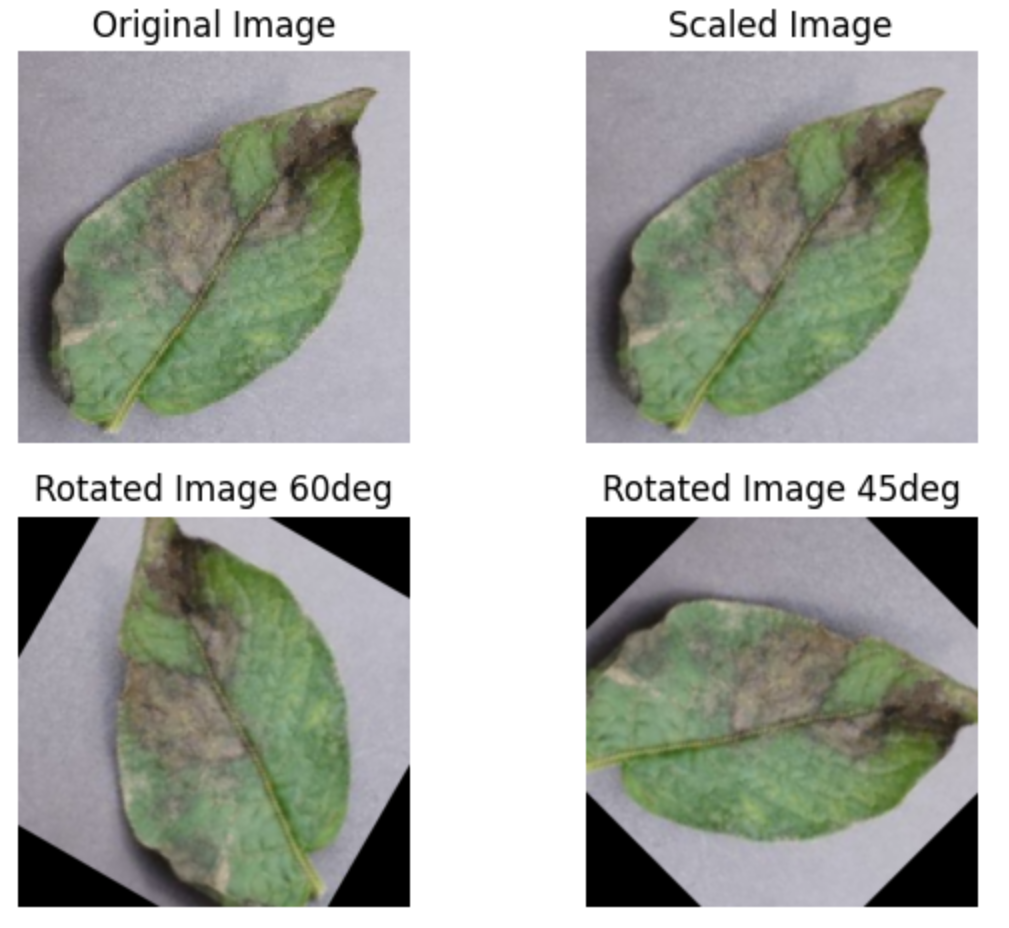
#### Detailed Features of the Dataset



#### Data Pre-processing

#### A close-up of a leaf Description automatically generated

**Fig.1** Sample Potato leaf



**Fig.2** Data Augmentation

From Figure 2, We applied various transformations to the images in our dataset. Firstly, we scaled

the images using a scaling factor to standardize their dimensions and ensure uniformity across the

dataset. We employed rotation transformations to generate variations of the original images,

simulating different viewing angles and perspectives.

A close-up of several leaves

Description automatically generated

**Fig.3** Interpolation

From Figure 3, We experimented with three commonly used interpolation methods: INTER\_LINEAR, INTER\_CUBIC, and INTER\_NEAREST. INTER\_LINEAR employs linear interpolation between neighboring pixels, INTER\_CUBIC utilizes cubic interpolation for smoother results, and INTER\_NEAREST performs nearest-neighbor interpolation by selecting the pixel value closest to the desired location.

A collage of images and graphs

Description automatically generated

**Fig .4** Graphical Representation

From Figure 4, We applied histogram equalization and contrast enhancement techniques to enhance the images. We generated histograms for both the original grayscale images and their equalized counterparts, revealing improvements in contrast and feature visibility.

#### Model Building

The method we used involved building a hybrid model that combined the strengths

of two powerful architectures, ResNet50 and Vision Transformer (ViT). First, we

used a ResNet50 model that had already been trained. This model is well-known for

its ability to extract complex characteristics from photos. ResNet50 was skilled at

capturing high-level representations necessary for picture understanding because it

had been trained on the large ImageNet dataset. We modified ResNet50 to fit our

classification goal by removing its top layers but keeping its convolutional

backbone, which allowed it to continue extracting features.

Following their extraction, a Vision Transformer (ViT) model was used to classify

the features. This part of our hybrid design consisted of a softmax activation layer

placed after a dense layer to enable multiclass classification. Our model combined

ViT's classification skills with ResNet50's feature extraction capabilities to create a

strong framework for identifying complex patterns and correctly classifying photos.

The training phase involved optimizing the hybrid model's parameters to minimize

classification errors and enhance its predictive performance. Using a labeled

dataset and the Adam optimizer in conjunction with the categorical cross-entropy

loss function, we trained the model. To avoid overfitting and guarantee the model's

capacity for generalization, we kept a careful eye on its performance on a

validation set during training. We sought to improve the model's accuracy and

discriminative capacity by iteratively modifying its parameters as it classified

photo into different categories.

Following training, we evaluated the hybrid model's performance on a separate test

dataset to gauge its accuracy and generalization ability. This assessment shed light

on how well the model predicted the future based on data that had never been seen

before. We evaluated the model's ability to distinguish between several classes and

correctly categorize photos by comparing its predictions to ground truth labels.

The resulting accuracy rate on the test dataset demonstrated how well our hybrid

model captured important features and used them to achieve accurate

classification.

A screenshot of a computer program

Description automatically generated

**Fig.5** Proposed Model Summary

#### Methodology of the system

A screenshot of a computer

Description automatically generated

**Fig.6** Model Architecture

1. Algorithm For Proposed Model

**Algorithm 1 Hybrid transformer With ResNet50**

**Import TensorFlowDefine**

**ResNet50 model**

*base\_model←tf.keras.applications.ResNet50*

*weights←’imagenet’input\_shape←(224,224,3)*

**Define Feature Extractor**feature\_extractor←tf.keras.Model

**Define Create Vision Transformer Classifier**

*inputs←tf.keras.Input(shape=(7,7,2048))*

*x←tf.keras.layers.Flatten()(inputs)*

*x←tf.keras.layers.Dense()*

*model←tf.keras.Model(inputs=inputs,outputs=outputs)*

**Define Hybrid Model**

*resnet\_input←tf.keras.Input(shape=(224,224,3))*

*resnet\_features←feature\_extractor(resnet\_input)*

*vit\_output←Create\_Vision\_Transformer\_Classifier()*

*model←tf.keras.Model(inputs=resnet\_input,outputs=vit\_output)*

**Load Datasets**

*train\_ds←tf.keras.utils.image\_dataset*

*test\_ds←tf.keras.utils.image\_dataset*

*validation\_ds←tf.keras.utils.image\_dataset*

**Compile and Train the Model**

*model.compile()*

*history←model.fit()*

**Evaluate the Model**

*test\_accuracy←model.evaluate(test\_ds)*

*"Test Accuracy:", test\_accuracy[1]​*

#### Model Evaluation

1. Comparative Results

A table with numbers and text

Description automatically generated

Our proposed model achieved an accuracy of 99.33% in detecting diseases in potato plant leaves. Compared to existing methods, it outperforms many, including YOLOv5 with ResNet50, which achieved 61.01% accuracy, and Region-Based Convolution Neural Network, which reached 93.08%. Notably, our model remains competitive with its high accuracy while focusing specifically on potato plant leaf disease detection.

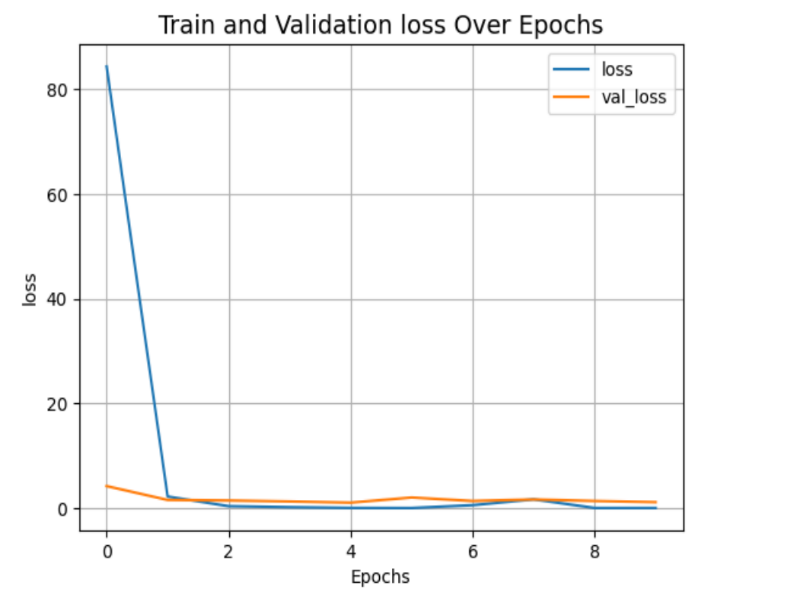
* Our Proposed Model

A close up of a white background

Description automatically generated

The ResNet50 model achieved a validation accuracy of 100% after 9 epochs, with a slight fluctuation in subsequent epochs. The Vision Transformer model attained an accuracy of 51.3% after 3 epochs, indicating slower convergence compared to ResNet50. In contrast, our proposed model achieved an impressive accuracy of 99.33% after 20 epochs, demonstrating superior performance and robustness in potato plant leaf disease detection.

* Train & Validation Plots

1. **Hybrid Transformer**

A graph with blue line and orange line

Description automatically generated

1. AUC & ROC Plots

The Receiver Operating Characteristic (ROC) curve illustrates the performance of a binary classification model across various threshold settings. It plots the true positive rate (sensitivity) against the false positive rate (1 - specificity).

The Area Under the ROC Curve (AUC) quantifies the model's ability to distinguish between the classes, with a higher AUC indicating better discrimination.

ROC and AUC are commonly used to evaluate and compare the performance of classification models, providing insights into their ability to balance true positives and false positives across different threshold values.

A graph with a line and numbers

Description automatically generated with medium confidence

1. Classification Report (Recall, Precision, Accuracy formulas)

A screenshot of a computer screen

Description automatically generated

**Accuracy: Accuracy** measures the proportion of correctly classified instances out of the total instances. It is calculated as:

**Precision:** Precision measures the proportion of true positive predictions out of all positive predictions. It is calculated as:

**Recall (Sensitivity):** Recall measures the proportion of true positive predictions out of all actual positive instances. It is calculated as:

**F1 Score:** F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is calculated as:

These metrics are commonly used to evaluate the performance of classification models, providing insights into their accuracy, ability to identify relevant instances (precision), ability to capture all relevant instances (recall), and overall balance between precision and recall (F1 score).True Positive is named as TP, True Negative(TN),False Positive(FP),False Negative(FN).

**Binary Cross Entropy Loss Function:**

Binary cross-entropy loss is a common loss function used in binary classification tasks, measuring the difference between predicted probabilities and actual binary labels. It penalizes predictions that deviate from the true labels, with higher penalties for larger discrepancies. The formula for binary cross-entropy loss is:

Where,

is the number of samples ,

is the true labels

is the predicted probablity of class i.

#### 

**Conclusion**

By giving farmers vital information on the severity of the disease, this optimized

approach has a lot of potential for real-world applications in agriculture. This method

enables farmers to take proactive steps to safeguard crop health and improve

agricultural sustainability by precisely identifying plant diseases early on. The study's

findings highlight how crucial it is to use deep learning models to overcome significant

obstacles in plant disease identification, as doing so will ultimately improve agricultural

techniques and crop manag ement plans.

Research on the creation and assessment of a hybrid model for identifying leaf diseases

in potato plants is the main topic of this work. The research made notable progress in

disease detection accuracy by merging the qualities of the Vision Transformer, which

is skilled at managing picture data, and ResNet50, which is renowned for its image

recognition capabilities. More precisely, the combination of these models produced an

astounding accuracy of 99.3%, even though ResNet50 and the Vision Transformer each

had an accuracy of 98.3% and 51.3%, respectively.

**Acknowledgment**

I would like to express my gratitude to my research advisor and team for their guidance and hard work in completing my research paper "Potato Leaf Disease Detection using Hybrid Transformer with RestNet50". I would also like to thank the participants of the research study for their valuable contribution.

### REFERENCES

1. Hasibul islam peyal , Md.Nahiduzzaman , Md. Abu Hanif Pramanik , Md. Khaild Syfullah,saleh Mohammed Shahriar , Abida Sultana, Mominul Ashan , Julfikar Haider , “Plant Disease Classifier: Detection of Dual-Crop Diseases Using Lightweight 2D CNN Architecture”.
2. Wassawa Shafik ,Ali Tufail,Abdallah Namoun,Liyanage Chandratilak de dilva,Rosyzie Anna Awg Haji Mohd Apong “A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends”.
3. R.Satya Rajendra Singh ,Rakesh Kumar Sanodiya “Zero-Shot Transfer Learning Framework for Plant Leaf Disease Classification”.
4. Khalid M.Hosny ,Walaa M.El-Hady,Farid M.Samy ,Eleni Vrochidou ,George A.Papakostas “Multi-Class Classification of Plant Leaf Diseases Using Feature Fusion of Deep Convolutional Neural Network and Local Binary Pattern”.
5. Md.Sakib Hossain Shovon, Shakrin Jahan Mozumder,Osim Kumar Pal,M.F Mridha,Nobuyoshi Asai ,Jungpil Shin “PlantDet: A Robust Multi-Model Ensemble Method Based on Deep Learning For Plant Disease Detection”.
6. Diana Susan Joseph , Pranav M.Pawar , Kaustubh Chakradeo “Real-Time Plant Disease Dataset Development and Detection of Plant Disease Using Deep Learning”.
7. Xin Zhang,Yuxin Mao,Qi Yang , Xuyang Zhang , “A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network”.
8. Vasileios Balafas,Emmanouil Karantoumains,Malamti Louta,Nikolaos Ploskas “Machine Learning and Deep Learning for Plant Disease Classification and Detecttion”.
9. Serosh Karim Noon, Muhammad Amjad,Muhammad Ali Qureshi,Abdul Mannan “Handling Severity Levels of Multiple Co-Occurring Cotton Plant Diseases Using Improved YOLOX Model”.
10. Muhammad Hammad Saleem , Johan Potgieter, Khalid Mahmood Arif, “A Performance-Optimized Deep Learning-Based Plant Disease Detection Approach for Horticultural Crops of New Zealand”.
11. Chunduri Madhurya,Emerson Ajith Jubilson, “YR2S: Efficient Deep Learning Technique for Detecting and Classifying Plant Leaf Diseases”.
12. Dataset from kaggle , Potato Leaf Disease Dataset Muhammad Ardi Putra : https://www.kaggle.com/datasets/muhammadardiputra/potato-leaf-disease-dataset