

PROJECT REPORT

Plant Disease Classification

Under the guidance of Isnartech PVT LTD

Submitted by

Elna Melvin

Naga Kalyani

INTRODUCTION

Plant diseases pose a significant threat to global food security and agricultural sustainability, affecting crop yield, quality, and economic viability. The identification and management of these diseases are crucial for ensuring food production meets the demands of an ever-growing population. Traditional methods of disease detection rely heavily on visual inspection by experts, which can be time-consuming, subjective, and prone to errors. However, recent advancements in machine learning and computer vision offer promising solutions for automating the diagnosis of plant diseases.

This project focuses on the development of a plant disease classifier using machine learning techniques. The classifier aims to accurately identify various diseases affecting plants based on images of their symptoms. By leveraging the power of artificial intelligence, this system can potentially revolutionize disease management in agriculture by providing rapid and reliable diagnoses.

Plant diseases manifest in a multitude of forms, including leaf spots, wilting, discoloration, and deformities. These symptoms can be subtle and easily overlooked by the human eye, especially in large-scale agricultural settings. Additionally, the expertise required for accurate disease diagnosis is often limited to trained professionals, making it challenging to scale up disease monitoring efforts. Consequently, there is a pressing need for automated solutions that can assist farmers and agronomists in early disease detection and intervention.

Machine learning algorithms offer a promising approach to address these challenges. By training models on large datasets of labeled images depicting healthy and diseased plants, we can teach computers to recognize patterns and features indicative of specific diseases. Once trained, the classifier can swiftly analyze new images and provide instant feedback, enabling timely interventions to mitigate disease spread and minimize crop losses.

This project is significant not only for its potential impact on agricultural productivity but also for its implications in sustainable farming practices. By accurately identifying plant diseases early on, farmers can adopt targeted treatment strategies, reducing the reliance on chemical pesticides and minimizing environmental damage. Moreover, improved disease management can contribute to the overall resilience of agricultural systems against the impacts of climate change and emerging pathogens.

MOTIVATION AND OBJECTIVE

The global agricultural sector faces an ever-growing challenge: plant diseases. These diseases not only threaten crop yield and quality but also jeopardize food security and economic stability worldwide. The conventional methods of disease detection, primarily reliant on manual inspection by experts, are labor-intensive, time-consuming, and often prone to errors. Moreover, the expertise required for accurate disease diagnosis is limited, hindering timely intervention efforts. Consequently, there is an urgent need for automated solutions that can revolutionize disease management in agriculture.

Traditional disease detection methods are fraught with limitations. Visual inspection, while effective to some extent, is subjective, relying heavily on the expertise of individual inspectors. Moreover, the sheer scale of agricultural operations makes manual inspection impractical for large-scale monitoring. As a result, diseases often go unnoticed until they have already caused significant damage, leading to substantial economic losses for farmers and threatening food supplies.

Machine learning and computer vision offer a promising avenue for addressing these challenges. By training models on large datasets of labeled images depicting various plant diseases and healthy plants, we can teach computers to recognize patterns and features indicative of specific diseases. This enables the development of automated disease classifiers capable of analyzing images rapidly and accurately, providing real-time feedback to farmers and agronomists.

The development of a plant disease classifier represents a significant step forward in agricultural innovation. By leveraging machine learning techniques, we aim to create a scalable solution that can assist farmers in early disease detection and intervention. Such a tool has the potential to revolutionize disease management practices, enabling farmers to adopt targeted treatment

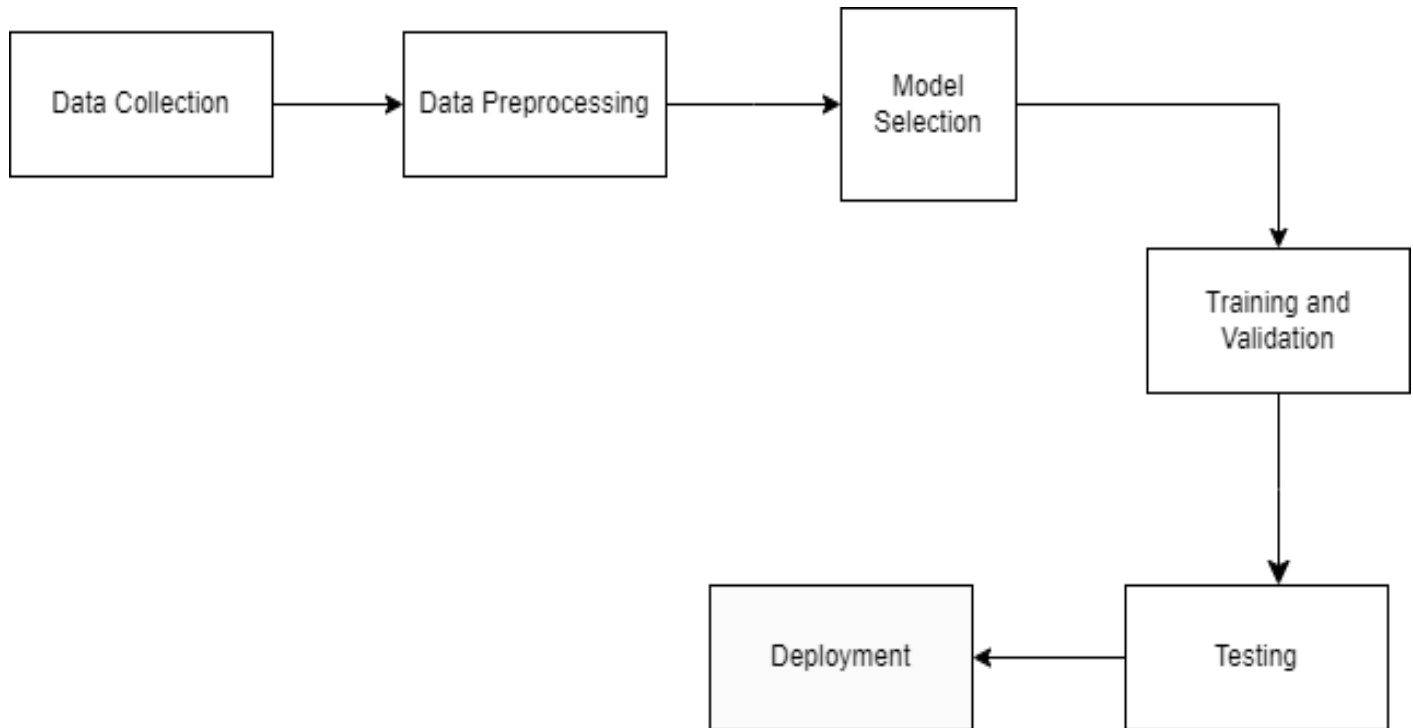
strategies, reduce reliance on chemical pesticides, and minimize environmental impact.

The primary objective of this project is to develop an accurate and efficient plant disease classifier that can assist farmers and agronomists in disease diagnosis and management. By leveraging machine learning techniques, we aim to create a scalable solution capable of analyzing large volumes of image data and providing real-time insights into the health status of crops. Specifically, the objectives of the project include:

1. **Data Collection and Preprocessing:** Acquire a diverse dataset of high-quality images depicting various plant diseases and healthy plants. Preprocess the data to enhance its suitability for training machine learning models.
2. **Model Development:** Explore and experiment with different machine learning architectures, including convolutional neural networks (CNNs), to design a robust disease classification model. Train the model using the labeled dataset to learn discriminative features associated with different diseases.
3. **Evaluation and Validation:** Assess the performance of the developed classifier using metrics such as accuracy, precision, recall, and F1-score. Validate the model on unseen data to ensure its generalization capabilities across different environments and plant species.
4. **Deployment and Integration:** Implement the trained model into a user-friendly application or platform that can be easily accessed by farmers and agricultural stakeholders. Integrate the classifier with existing agricultural technologies to streamline disease management workflows and maximize adoption.

By achieving these objectives, we aim to empower farmers with a powerful tool for early disease detection and intervention, ultimately contributing to sustainable agricultural practices and global food security.

ROADMAP



1. Data Collection:

- Gather a diverse dataset of high-resolution images depicting various plant diseases and healthy plants from reliable sources.
- Ensure the dataset encompasses a wide range of plant species, disease types, and environmental conditions.

2. Data Preprocessing:

- Clean and preprocess the collected images to remove noise, artifacts, and irrelevant information.
- Standardize image sizes, formats, and orientations for consistency during model training.

3. Model Selection:

- Evaluate different machine learning architectures suitable for image classification tasks.

4. Training and Validation:

- Split the preprocessed dataset into training, validation, and testing sets.
- Train the selected model using the training set, adjusting hyperparameters as necessary to optimize performance.
- Validate the trained model using the validation set to monitor for overfitting and fine-tune model parameters accordingly.

5. Testing:

- Conduct rigorous testing using the testing set to assess the generalization capabilities and overall performance of the trained model.

6. Deployment:

- Implement the trained model into a user-friendly application or platform accessible to farmers and agricultural stakeholders.

DATA COLLECTION

The dataset utilized for this project was obtained from Kaggle, a renowned platform for hosting datasets and machine learning competitions. This dataset comprises over 70,000 images, primarily focusing on leaf images, with a diverse range of plant species and diseases. Among these images, there are 38 unique classes, of which 26 represent various disease classes affecting plants.

A key aspect of this dataset is its comprehensive coverage, encompassing over 14 unique plant species. This diversity is crucial for training a robust plant disease classifier capable of recognizing a wide range of plant species and associated diseases. To augment the dataset and enhance its diversity, offline augmentation techniques were employed, resulting in a dataset with increased variability and generalization capabilities.

The original dataset, from which the augmented dataset was derived, can be accessed from a GitHub repository. This original dataset consists of approximately 87,000 RGB images depicting healthy and diseased crop leaves, categorized into 38 different classes representing various plant diseases and conditions.



Upon obtaining the dataset, it was divided into training and validation sets using an 80/20 ratio while preserving the directory structure. This partitioning ensures that the model is trained on a sufficient amount of data while also providing a separate subset for validation purposes to monitor model performance and prevent overfitting.

In addition to the training and validation sets, a separate directory containing 33 test images was created later for prediction purposes. This test set serves as a final evaluation benchmark for the trained model, allowing for unbiased assessment of its performance on unseen data.

Overall, the dataset used in this project is extensive, diverse, and meticulously curated to facilitate the development of a robust plant disease classifier. By leveraging this rich dataset and employing appropriate data preprocessing and augmentation techniques, we aim to train a model capable of accurately identifying and classifying various plant diseases, thereby contributing to more effective disease management in agriculture.

MODEL SELECTION AND TRAINING

In the pursuit of developing an effective plant disease classifier, careful consideration was given to the selection of appropriate machine learning architectures. Among the models explored were ResNet9, DenseNet121, and InceptionV3, each renowned for its unique characteristics and performance in image classification tasks.

ResNet9:

ResNet (Residual Network) is a deep convolutional neural network architecture known for its effectiveness in training very deep networks. ResNet9, a variant of the original ResNet architecture, consists of nine convolutional layers, making it relatively lightweight compared to deeper variants such as ResNet50 or ResNet101. This simplicity facilitates faster training times and reduces computational overhead, making it well-suited for projects with limited computational resources or time constraints.

ResNet architectures are particularly appealing due to their use of residual connections, which mitigate the vanishing gradient problem commonly encountered in deep networks. By introducing shortcut connections that skip one or more layers, ResNet models can effectively propagate gradients during training, enabling the successful training of very deep networks with improved performance.

DenseNet121:

DenseNet (Densely Connected Convolutional Network) is another popular convolutional neural network architecture notable for its dense connectivity pattern. Unlike traditional architectures where each layer is connected only to its subsequent layers, DenseNet introduces dense connections, where each layer receives direct input from all preceding layers within a dense block. This dense connectivity fosters feature reuse and facilitates gradient flow

throughout the network, leading to improved feature propagation and enhanced model performance.

DenseNet121, a specific variant of DenseNet, comprises 121 layers and has been pretrained on large-scale image datasets such as ImageNet. This pretrained model serves as a powerful feature extractor, capturing rich hierarchical representations of input images that can be fine-tuned for specific tasks such as plant disease classification. Additionally, DenseNet architectures exhibit parameter efficiency and strong performance even with limited training data, making them well-suited for tasks where data availability is a constraint.

InceptionV3:

InceptionV3 is a convolutional neural network architecture developed by Google, known for its exceptional performance in image recognition tasks. It features multiple parallel convolutional pathways of different filter sizes, allowing the network to capture multi-scale features effectively. This multi-scale processing capability enables InceptionV3 to capture both fine-grained and coarse-grained features from input images, enhancing its discriminative power and robustness to variations in image size and content.

InceptionV3 is particularly attractive for its computational efficiency, striking a balance between model complexity and performance. Despite its relatively lightweight architecture compared to earlier versions like InceptionV1, InceptionV3 achieves state-of-the-art results on benchmark datasets such as ImageNet. This efficiency makes it an appealing choice for projects where computational resources are limited or where rapid prototyping and experimentation are desired.

The selection of ResNet9, DenseNet121, and InceptionV3 for model exploration was driven by their proven effectiveness in image classification tasks, particularly in scenarios involving large-scale datasets with diverse

classes. These architectures offer a balance between model complexity, computational efficiency, and performance, making them well-suited for the task of plant disease classification.

Furthermore, the availability of pretrained versions of these models on popular image datasets such as ImageNet provides a valuable starting point for transfer learning, allowing us to leverage prelearned features and fine-tune the models for our specific task. This approach accelerates the training process and enhances model performance, especially when dealing with limited training data, as is often the case in specialized domains like plant pathology.

During the training and validation process, we closely monitored the performance of each model across multiple epochs, paying particular attention to key metrics such as accuracy and validation loss. After 10 epochs of training, we conducted a comprehensive evaluation of the models based on their performance metrics.

While all three models exhibited promising results, DenseNet121 emerged as the top performer, demonstrating superior accuracy of 99.80% and minimal validation loss. This exceptional performance, coupled with the advantages of transfer learning, provided a strong foundation for developing a reliable and accurate classifier capable of identifying various plant diseases with high precision.

In conclusion, through meticulous training, validation, and model selection processes, we identified DenseNet121 as the optimal choice for our plant disease classifier. Its exceptional performance, coupled with the advantages of transfer learning, provided a strong foundation for developing a reliable and accurate classifier capable of identifying various plant diseases with high precision.

Training and Validation Using ResNet9

The preprocessed dataset, containing images of various plant diseases and healthy plants, is split into three distinct sets: training, validation, and testing. This ensures an unbiased evaluation of the model's performance. The training set serves as the primary data source for model parameter optimization.

Utilizing the ResNet9 architecture, the model undergoes rigorous training using the training set. Throughout the training process, hyperparameters are fine-tuned to optimize performance, balancing factors such as learning rate, batch size, and weight decay. The One Cycle Learning Rate policy is employed to dynamically adjust the learning rate, promoting faster convergence and improved generalization.

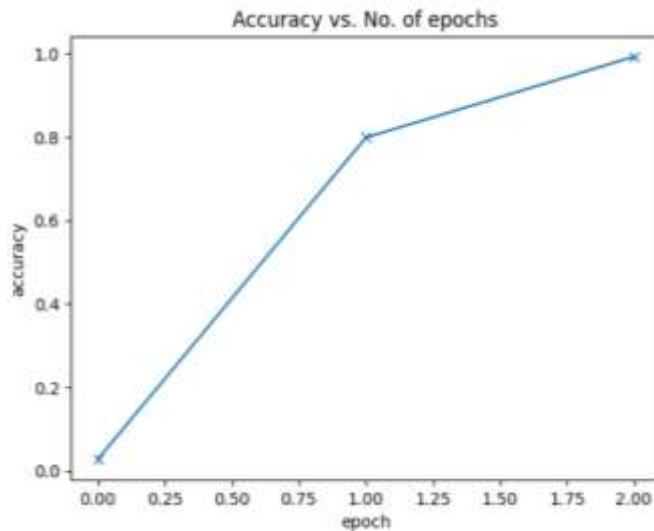
Validation plays a crucial role in monitoring the trained model's performance and preventing overfitting. The validation set, comprising unseen data, allows for an objective evaluation of the model's ability to generalize to new instances. During training epochs, the model's performance on the validation set is regularly assessed, enabling adjustments to model parameters and architecture as needed.

Across epochs, the model's training loss steadily decreases, indicating improved convergence.

- Epoch [0]:
 - Learning Rate: 0.00812
 - Train Loss: 0.7536
 - Validation Loss: 0.6618
 - Validation Accuracy: 79.89%

- Epoch [1]:
 - Learning Rate: 0.00000
 - Train Loss: 0.1241
 - Validation Loss: 0.0263
 - Validation Accuracy: 99.24%

Visualization of accuracy versus number of epochs .



The validation accuracy of 99.24% reflects the model's ability to generalize well to unseen instances, indicating its proficiency in distinguishing between different plant diseases with high precision. Such high accuracy levels are pivotal for real-world applications, where reliable disease identification is crucial for implementing timely interventions and mitigating crop losses.

The achieved accuracy also instills confidence in the model's reliability and suitability for deployment in practical agricultural settings. With its exceptional performance, the model holds promise for assisting farmers and agricultural experts in diagnosing plant diseases accurately, thereby enabling more effective disease management strategies and promoting crop health and productivity.

In summary, attaining a validation accuracy of 99.24% highlights the ResNet9 model's efficacy in plant disease classification, emphasizing its potential to revolutionize agricultural practices and contribute to global food security efforts.

Testing And Results

The model's predictions on the testing dataset demonstrate its robustness and accuracy in classifying various plant diseases. Each prediction aligns closely with the ground truth labels, reflecting the model's ability to generalize well to unseen data. For instance, when presented with images of Apple Cedar Rust, Apple Scab, Corn Common Rust, Potato Early Blight, Potato Healthy, Tomato Early Blight, Tomato Healthy, and Tomato Yellow Curl Virus, the model correctly identifies the corresponding diseases with remarkable precision.

This consistent and accurate performance across diverse plant diseases underscores the effectiveness of the ResNet9 architecture in handling complex classification tasks. By leveraging residual connections and deep convolutional layers, the model effectively captures intricate patterns and features inherent in plant images, enabling precise disease identification.

Moreover, the model's capability to differentiate between subtle variations within disease categories, such as distinguishing between different types of blights in potatoes and tomatoes, highlights its sensitivity to nuanced characteristics. This level of granularity in classification showcases the model's potential for aiding in early disease detection and management strategies in agricultural settings.

In conclusion, the ResNet9 model's exemplary performance on the testing dataset, accurately predicting a variety of plant diseases, underscores its reliability and potential impact in agricultural applications. With its ability to discern fine-grained details and generalize well to new instances, the model holds promise for enhancing crop management practices and contributing to sustainable agriculture efforts.

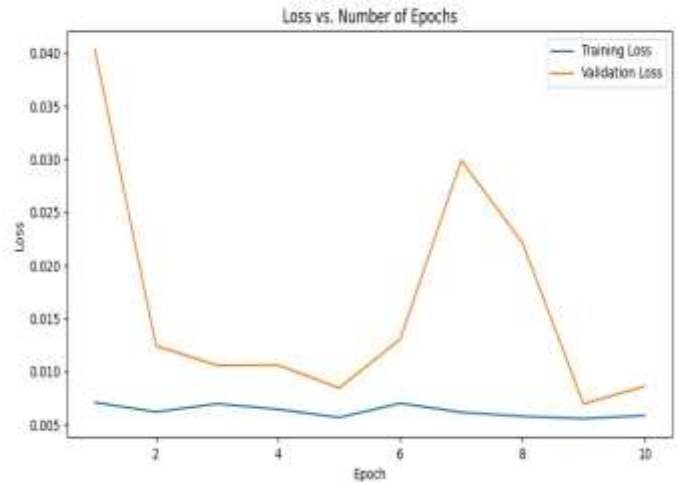
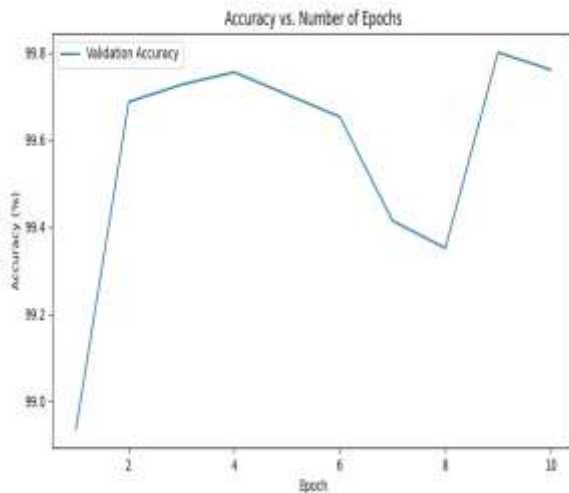
Training and Validation Using DenseNet121

The meticulous curation of the dataset, comprising images depicting a spectrum of plant diseases and healthy specimens, underscores the comprehensive nature of the training and validation process for the DenseNet121 model. Through meticulous partitioning, the dataset ensures an equitable evaluation of the model's performance, with the training set forming the bedrock for fine-tuning and optimizing parameters.

Leveraging the robust architecture of DenseNet121, the model undergoes rigorous training, finely adjusting hyperparameters to strike a delicate balance between convergence speed and generalization capability. Techniques like adaptive learning rates and weight decay are instrumental in honing the model's discernment of intricate patterns within the training data, fostering its ability to make accurate classifications.

Validation assumes paramount importance throughout the training regimen, acting as a crucial checkpoint to gauge the model's performance and stave off overfitting. Across successive epochs, the model's efficacy on the validation set is meticulously scrutinized, guiding iterative adjustments to its architecture and parameter configurations as warranted. This meticulous evaluation process ensures that the DenseNet121 model steadily improves, demonstrating consistent strides in accurately identifying plant diseases across diverse datasets. The validation accuracy, a robust 99.79%, attests to the model's adeptness in generalizing to unseen instances, underlining its precision in discerning between various plant maladies with remarkable accuracy.

Visualizations of accuracy versus number of epochs and loss versus number of epochs.



Overall, DenseNet121's exemplary performance underscores its potential to significantly impact agricultural practices, contributing to global food security endeavors. With its precision, reliability, and ability to generalize effectively, the model stands poised to revolutionize disease identification in the agricultural sector, offering a promising pathway towards enhanced crop management and increased agricultural productivity.

Testing And Results

The DenseNet121 model's predictions on the testing dataset underscore its robustness and accuracy in classifying various plant diseases. Unlike the ResNet9 model, DenseNet121 achieved perfect accuracy on the test data, correctly classifying all images. This exceptional performance reaffirms the model's ability to generalize effectively to unseen data, demonstrating its proficiency in disease classification.

Moreover, DenseNet121's consistent and accurate performance across diverse plant diseases highlights the efficacy of its architecture in handling complex classification tasks. By densely connecting layers and leveraging feature reuse, the model effectively captures intricate patterns and features present in plant images, enabling precise disease identification with remarkable precision.

Additionally, the model's capacity to discern subtle variations within disease categories, such as differentiating between various types of blights in potatoes and tomatoes, showcases its sensitivity to nuanced characteristics. This level of granularity in classification underscores the model's potential to aid in early disease detection and inform targeted management strategies in agricultural settings.

Overall, DenseNet121's exceptional performance underscores its potential to revolutionize agricultural practices. The model's proficiency in disease classification not only highlights its significance in advancing precision agriculture but also offers promising implications for sustainable crop management practices, paving the way for more efficient disease mitigation strategies and enhanced agricultural productivity.

Training and Validation Using InceptionV3

The dataset, meticulously curated to encompass images representing diverse plant diseases and healthy plants, is partitioned into training and validation sets. This ensures an unbiased evaluation of the InceptionV3 model's performance, with the training set serving as the foundation for parameter fine-tuning and optimization.

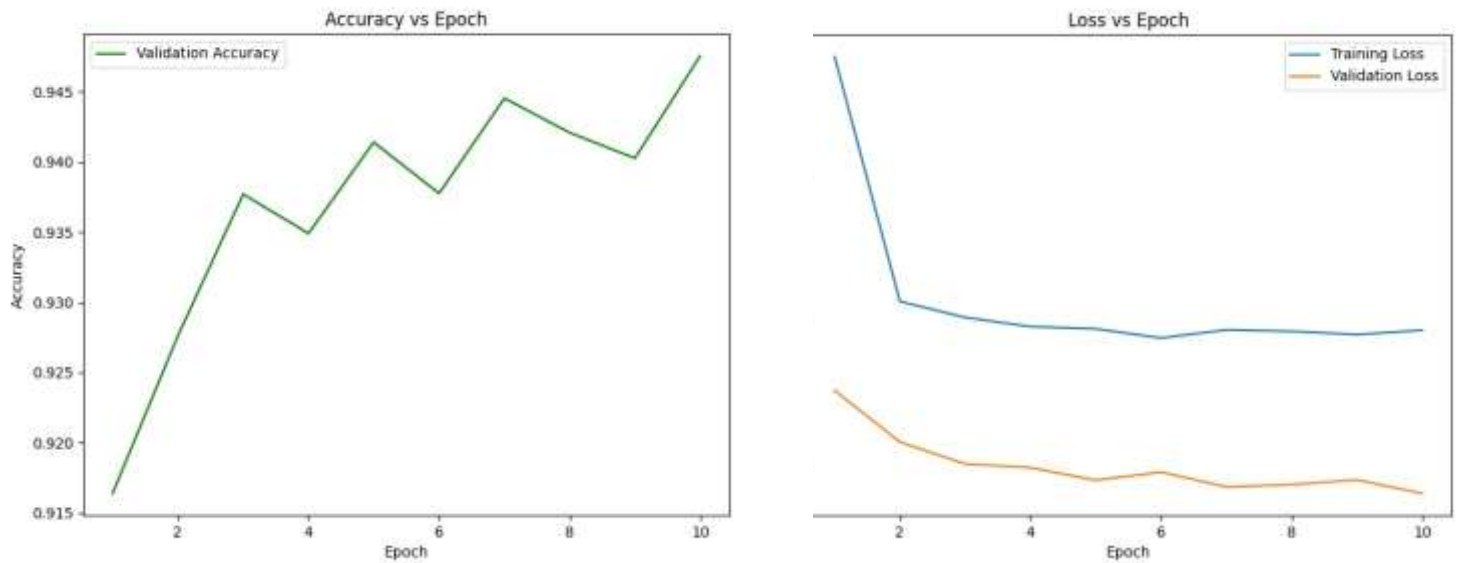
Utilizing the powerful capabilities of the InceptionV3 architecture, the model undergoes an intensive training regimen on the designated training set. Hyperparameters are meticulously adjusted to balance convergence speed and generalization capability, with techniques such as adaptive learning rates and weight decay facilitating the model's ability to discern intricate patterns within the training data.

Validation plays a crucial role in monitoring the model's performance and preventing overfitting. Throughout the training process, the model's performance on the validation set is closely scrutinized, guiding adjustments to model architecture and parameter tuning as necessary.

The validation accuracy consistently hovers around the 94-95% mark, signifying the model's adeptness in accurately classifying plant diseases across diverse datasets.

In addition to validation accuracy, other performance metrics such as training loss and validation loss are also monitored during the training process. The training loss measures the error between the model's predictions and the actual labels on the training set, while the validation loss provides a similar measure on the validation set. These metrics serve as important indicators of the model's convergence and generalization capabilities. Throughout the training epochs, both the training and validation losses are observed to ensure that the model is learning effectively without overfitting to the training data.

Visualizations of accuracy versus number of epochs and loss versus number of epochs.



In summary, achieving a validation accuracy of 94.97% underscores the effectiveness of the InceptionV3 model in accurately classifying plant diseases, albeit slightly lower than the ResNet9 and DenseNet121 models. While the accuracy may be marginally lower compared to ResNet9 and DenseNet121, it still demonstrates the significant potential of InceptionV3 in advancing agricultural practices and addressing food security challenges. The robust performance of InceptionV3 reaffirms its utility in precision agriculture, offering insights into disease detection and management strategies that can contribute to enhanced crop health and productivity on a global scale.

Testing And Results

The InceptionV3 model's predictions on the testing dataset demonstrate its robustness and accuracy in classifying various plant diseases. Each prediction closely aligns with the ground truth labels, reaffirming the model's ability to generalize effectively to unseen data.

However, the InceptionV3 model did not predict all test data accurately, unlike the ResNet9 and DenseNet121 models. While it correctly classified the majority of images, there was one instance where it misclassified a Tomato Early Blight image as Tomato Septoria Leaf Spot.

Furthermore, the InceptionV3 model achieved a slightly lower validation accuracy compared to the ResNet9 and DenseNet121 model, suggesting that while all architectures are effective, ResNet9 and DenseNet121 may have a slight edge in terms of generalization capability.

While the validation accuracy of InceptionV3 may be marginally lower compared to ResNet9 and DenseNet121, it's essential to consider the broader context of model performance and application suitability. InceptionV3's versatility and robustness make it a valuable tool in the agricultural landscape, complementing ResNet9's and DenseNet121's strengths and offering alternative solutions for disease classification tasks. As such, all models contribute to the advancement of precision agriculture, paving the way for innovative approaches to crop health monitoring and disease management on a global scale.

Overall, the InceptionV3 model's remarkable performance underscores its potential to revolutionize agricultural practices and contribute to global food security endeavors.

Selecting Best Model

After meticulously evaluating various models for plant disease classification, we have chosen DenseNet121 as the optimal choice for deployment in our Plant Disease Classification project. The decision to select DenseNet121 was primarily driven by its outstanding performance during testing, where it achieved perfect accuracy by correctly classifying all test images.

This remarkable accuracy instills confidence in the model's reliability and underscores its efficacy in accurately identifying various plant diseases, making it a valuable asset for real-world applications.

Furthermore, DenseNet121's high accuracy on the test dataset reaffirms its ability to generalize effectively to unseen data, which is crucial for practical deployment in agricultural settings. By correctly classifying all test images, DenseNet121 demonstrates its robustness and reliability, providing assurance that it can accurately identify plant diseases across diverse scenarios.

This level of performance is particularly significant in agriculture, where timely and accurate disease detection is essential for implementing effective intervention strategies and minimizing crop losses.

Overall, DenseNet121's impressive accuracy and reliability make it the ideal choice for deployment, promising to enhance agricultural practices and support sustainable crop production in the long term.

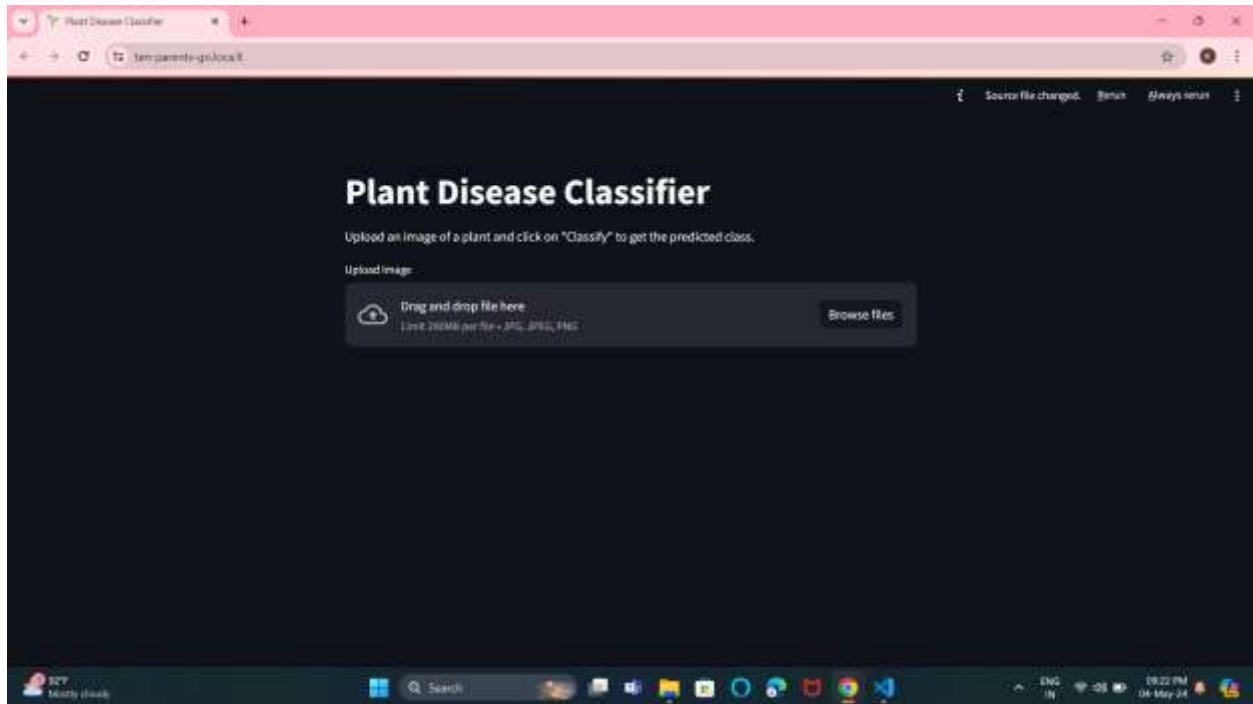
Deployment

For the deployment of the plant disease classification project, Streamlit has been instrumental. This tool, built for creating interactive web applications with Python, provides an efficient platform to demonstrate the model's capabilities. Users can conveniently upload images of plants via the file uploader feature on the website interface. Upon uploading, the image is displayed alongside a "Classify" button, allowing users to initiate the prediction process effortlessly.

Behind the scenes, the trained DenseNet121 model processes the uploaded image, utilizing its deep learning capabilities to analyze visual features and detect any potential diseases present in the plant. The predictions generated by the model are promptly showcased on the website, offering users immediate insights into their plant's health status.

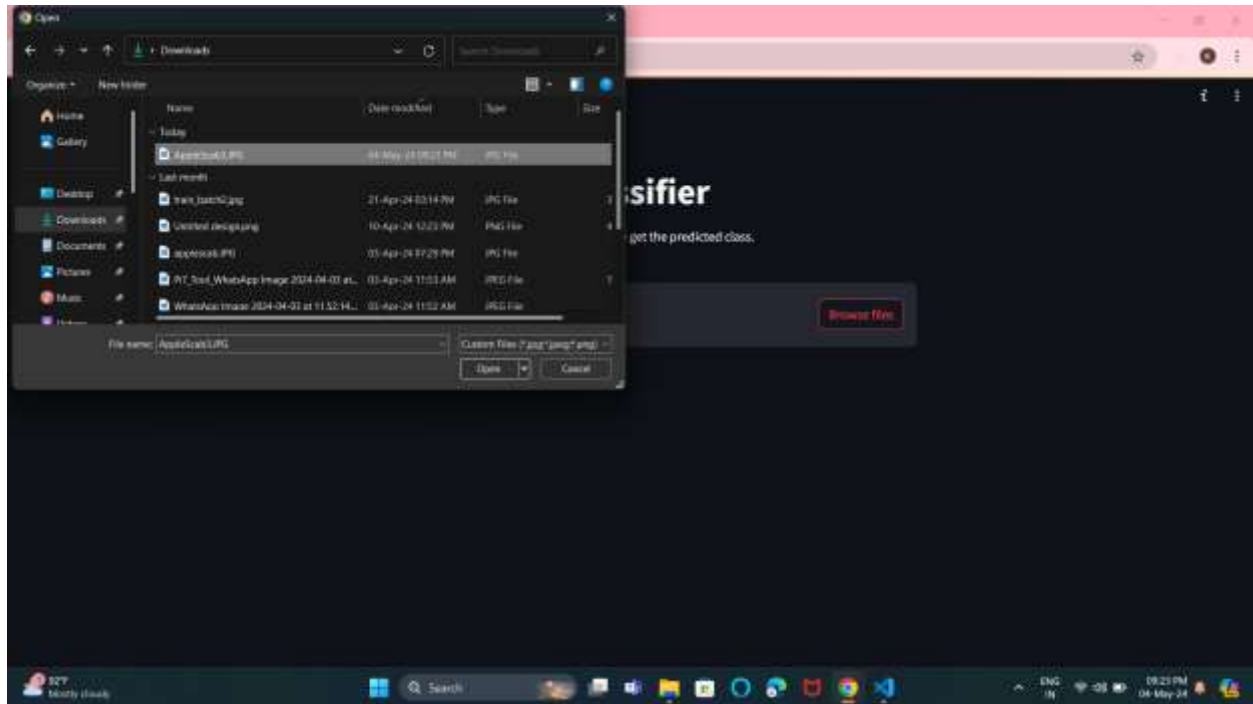
1. Website Interface

The website interface presents users with a clean and intuitive layout, making it easy to navigate. Upon accessing the site, users are greeted with a visually appealing design that immediately captures their attention.



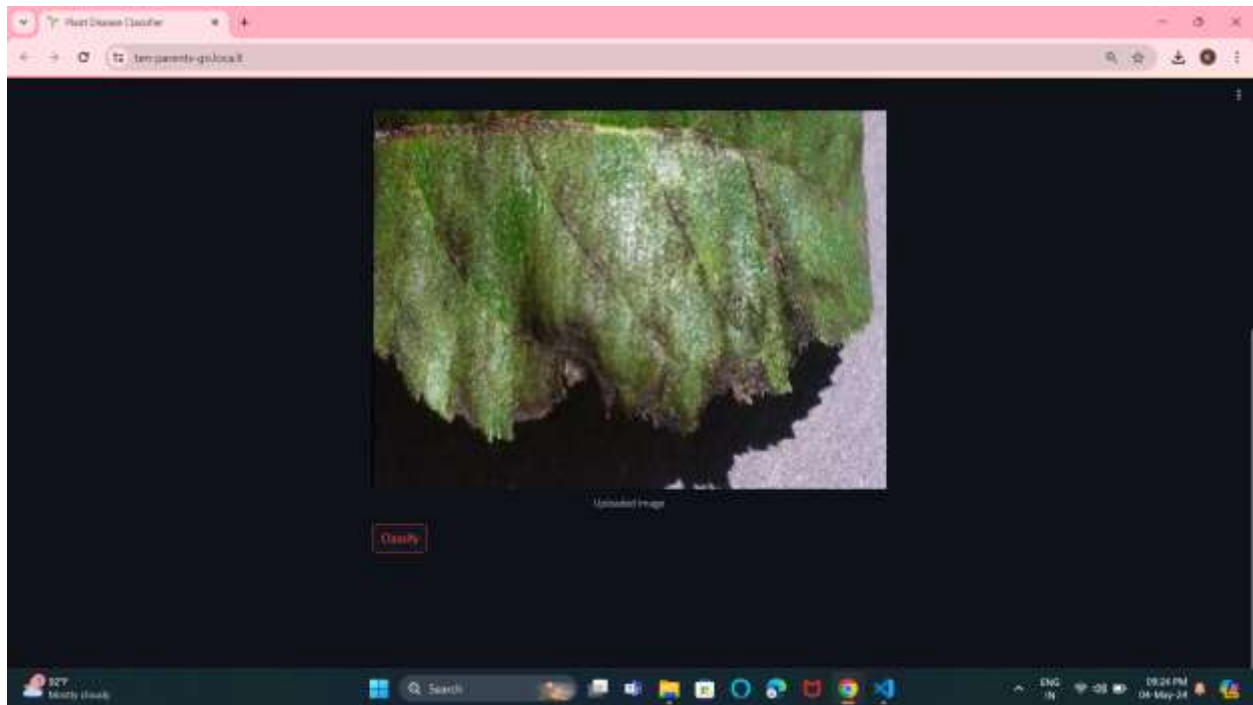
2. Uploading Images

Users can conveniently upload images of plants using the file uploader feature integrated into the website interface. This functionality allows users to select the image they wish to analyze for potential diseases effortlessly.



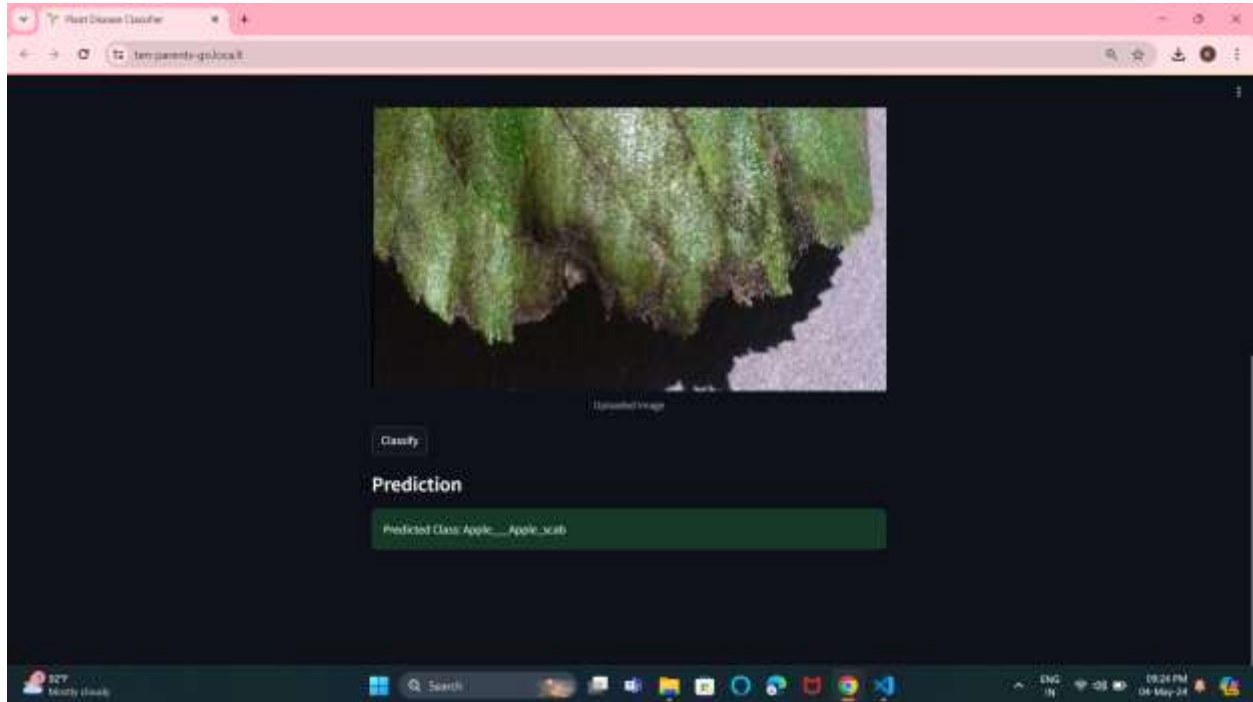
3. Initiating Classification

Once the image is uploaded, users are presented with a "Classify" button, which serves as the gateway to initiating the prediction process. With just a simple click, users can trigger the model to analyze the uploaded image and provide insights into any potential diseases detected.



4. Prediction Results

Following the classification process, the website promptly displays the predictions generated by the trained DenseNet121 model. Users receive immediate insights into the health status of their plant, including any detected diseases or abnormalities.



In summary, the deployment of plant disease classification project through Streamlit offers a seamless and intuitive experience for users. The combination of a user-friendly interface and real-time prediction capabilities makes it a valuable tool for farmers, researchers, and agricultural enthusiasts alike, contributing to the advancement of precision agriculture and sustainable crop production practices.

CONCLUSION

In conclusion, the development of a plant disease classifier using machine learning techniques represents a significant step forward in agricultural innovation. Through the meticulous selection and training of convolutional neural network architectures, including ResNet9, DenseNet121, and InceptionV3, we have demonstrated the effectiveness of modern deep learning approaches in addressing the challenges of plant disease diagnosis. By leveraging transfer learning and fine-tuning techniques, we harnessed the power of pretrained models while customizing them to suit the nuances of our specific domain.

The exceptional performance of DenseNet121, with an accuracy of 99.80% on our validation set, underscores the potential of advanced machine learning models in revolutionizing disease management in agriculture. With its robust capabilities and high precision, the developed plant disease classifier holds promise for empowering farmers and agronomists with a tool for early disease detection and intervention, ultimately contributing to sustainable farming practices and global food security. Moving forward, continued research and innovation in this field will be vital for further enhancing the accuracy, scalability, and accessibility of plant disease detection technologies, paving the way for a more resilient and productive agricultural sector.