ECC691: PROJECT I (B)

Professor: Dr. Dalia Nandi, IIIT Kalyani

Early Detection of Crop Diseases using Convolutional Neural Networks (CNN)

By:

Shirish Manoj Bobde
Reg: ECE/21152/812
Divyanshu Kumar
Reg: ECE/21123/783
Geetansh Jangid
Reg: ECE/21124/784

Bachelor Thesis submitted to Indian Institute of Information Technology, Kalyani for the partial fulfillment of the degree of Bachelor of Technology in Department of Electronics and Communication Engineering (Spring Semester 2023-2024)

Declaration

We hereby declare that the work being presented in this thesis entitled, "Early Detection of Crop Diseases using Convolutional Neural Networks (CNN)", submitted to Indian Institute of Information Technology, Kalyani in partial fulfillment for the award of the degree of Bachelor of Technology in Electronics and Communication Engineering during the period from January, 2023 to May, 2023 under the supervision of Dr. Dalia Nandi, Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Kalyani, West Bengal 741235, India, does not contain any information and any piece of code and text generated by AI enabled software.

Shirish Manoj Bobde – ECE/21152/812
Divyanshu Kumar – ECE/21123/783
Geetansh Jangid – ECE/21124/784

Department: Electronics and Communication Engineering
Institute: Indian Institute of Information Technology, Kalyani

This is to certify that the aforementioned statements made are correct to the best of my knowledge.

Signed:

Date:

Certificate

This is to certify that the thesis entitled "Early Detection of Crop Diseases using Convolutional Neural Networks (CNN)" being submitted by Shirish Manoj Bobde (ECE/21152/812), Divyanshu Kumar (ECE/21123/783) and Geetansh Jangid (ECE/21124/784) in the Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Kalyani, West Bengal, India, for the award of Bachelor of Technology in Electronics and Communication Engineering is an original research work carried by them under my supervision and guidance. The thesis has fulfilled all the requirements as per the regulations of Indian Institute of Information Technology, Kalyani and in my opinion, has reached the standards needed for submission. The work, techniques and the results presented have not been submitted to any other University or Institute for the award of any other degree or diploma.

Dr. Dalia Nandi
Assistant Professor

Department of Electronics and Communication Engineering Indian Institute of Information Technology, Kalyani

Date:

Acknowledgements

We would like to express our gratitude to Dr. Dalia Nandi, our supervisor and mentor, for her invaluable guidance and insightful feedback throughout the project. We are sincerely thankful for the opportunity she has provided us to undertake this challenging project. We eagerly anticipate furthering our understanding of this technology through this experience.

Shirish Manoj Bobde (ECE/21152/812)
Divyanshu Kumar (ECE/21123/783)
Geetansh Jangid (ECE/21124/784)

Department of Electronics and Communication Engineering Indian Institute of Information Technology, Kalyani

Date:

Contents

1 Overview

Our human society needs to increase the food production by approximately 70% by 2050 to feed and expected population of over 9 billion people, with 40% Loss in Crop Yield each year due to various factors including infectious disease contribute to the reduction of yield in developed nations whereas many farmers from the developing part of world experiencing yield loss as high as 100% . The widespread distribution of Smartphones and technology alike, among farmers and the world with an expected 5 billion smartphones by 2020 offers the potential of usability of smartphones as a valuable tool for farmers and alike. One of the many potential applications is the development of Mobile disease diagnostics through machine learning.[1]

Agricultural drones, also known as unmanned aerial vehicles (UAVs) have brought a transformation in modern agriculture by offering farmers and agronomists with a powerful tool for crop monitoring and management. Equipped with various sensors and imaging capabilities, these drones can capture high resolution images and data, providing valuable insights into crop health and environmental conditions. [9]

Phase I focuses on using Convolutional Neural Networks (CNNs) and deep learning to predict plant diseases, a critical task to mitigate losses in yield and maintain the quality. Plant diseases can cause economic losses and reduce food production, which emphasizes the importance of early detection and management. The project utilizes a dataset containing images of healthy plants and plants infected with various diseases. This dataset is needed for training the CNN model to recognize patterns and features indicative of different plant diseases. Data preprocessing techniques such as resizing, normalization, and augmentation are applied to ensure the uniformity and quality of the dataset. Next, for the Model Architecture, CNNs are well-suited for image classification tasks due to their ability to automatically learn spatial hierarchies of features from the images. The CNN architecture for this project consists of convolutional layers, followed by max-pooling layers to extract relevant features from the input images. These features are then flattened and passed through fully connected layers to make predictions about the presence of diseases. The model is trained using the preprocessed dataset to optimize its parameters for accurate disease classification. The training process involves minimizing a loss function using an Adam optimizer to update the model weights. The model's performance is monitored using a validation set to prevent overfitting. [4]

The successful development and deployment of this model can significantly impact agriculture by providing farmers with a tool for early detection and management of plant diseases. By identifying diseased plants early, farmers can take timely actions such as targeted treatments, thereby reducing the spread of diseases and minimizing crop losses.

2 Literature Review

In India, every year, 30% of crops are lost due to pests and diseases, resulting in an estimated annual loss of Rs. 90,000 crores. Increased use of pesticides due to late disease detection harms the environment and human health. Crop failure due to disease can lead to deforestation for new agricultural land.[1][3]

Visual inspection for disease identification in crops is a time-consuming, subjective process that requires expertise and is limited by coverage and scale. It involves regular field visits and close examination of plants, which can be impractical for busy farmers or large farms. Relying on visual inspection alone can lead to misdiagnosis, as different farmers may interpret symptoms differently. Additionally, early disease detection demands expertise and training, posing a challenge for inexperienced farmers. By the time visual symptoms are apparent, the disease may have already spread, reducing the effectiveness of control measures. Moreover, visual inspection's efficacy is restricted by farmers' capacity to physically examine all plants in a field, potentially leading to oversight in some areas. Scaling visual inspection for large agricultural operations or in challenging terrain presents significant challenges. The effectiveness of visual inspection is also subject to environmental factors like weather, lighting, and plant growth stage. Poor visibility or challenging environmental conditions can impede accurate disease identification.[5]

In the field of crop disease detection using Unmanned Aerial Vehicles (UAVs), several works have been undertaken to leverage UAV capabilities for efficient and timely crop health monitoring. Some key advancements and research areas include the use of multispectral and hyperspectral cameras on UAVs for detailed spectral information, enabling early detection of stress, diseases, and nutrient deficiencies based on spectral signatures. Additionally, remote sensing techniques utilizing Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Chlorophyll Index (CI) derived from UAV imagery are employed to assess crop health and identify disease hotspots.[9][4]

Machine learning algorithms such as Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) are integrated to analyze UAV-collected data and classify crop health conditions, including disease presence and severity. Furthermore, efforts are made to develop systems for onboard processing or transmitting UAV data for immediate analysis, enabling timely corrective actions in response to disease outbreaks.[11]

3 Methodology

The project aims to revolutionize crop disease detection through advanced machine learning techniques, specifically Convolutional Neural Networks (CNNs), in two key phases: single crop disease detection and multi-crop disease detection. These phases involve meticulous dataset collection, preprocessing, model selection, training, and evaluation to develop highly accurate models for detecting diseases in crops. By leveraging different CNN architectures such as ResNet and Inception for single crop disease detection and a versatile Keras model for multi-crop disease detection, the project aims to enhance crop management practices, leading to reduced yield losses and improved agricultural productivity.

In the **single crop disease detection phase**, the project collected a dataset comprising images of a single crop, including healthy and diseased plants, ensuring a diverse representation of diseases. Preprocessing steps such as resizing, normalization, and augmentation were performed to standardize image sizes, enhance quality, and ensure dataset consistency. Various CNN architectures were evaluated, with ResNet emerging as the most accurate model for single crop disease detection.[9] Transfer learning was applied by leveraging features learned from pre-trained models on large datasets like ImageNet to optimize model performance. The models were trained on the preprocessed dataset, fine-tuning the last layers for the specific task of single crop disease detection. Model performance was evaluated using standard metrics like accuracy, precision, recall, and F1 score.

In the **multi-crop disease detection phase**, the dataset was expanded to include images of various crops, each with its unique disease characteristics, ensuring a diverse representation. Similar preprocessing steps as in single crop detection were applied to maintain consistency and quality across different crop types. A Keras model was selected for its flexibility and ease of implementation in multi-crop disease detection. The Keras model was trained on the expanded dataset, adjusting hyperparameters as necessary for optimal performance. Model evaluation using standard metrics validated its ability to detect diseases across different crop types. While ResNet demonstrated superior accuracy in single crop disease detection, the Keras model showed comparable performance across multiple crops, highlighting its effectiveness and versatility.

This project represents a significant advancement in the field of crop disease detection, leveraging Convolutional Neural Networks (CNNs) to enhance agricultural practices. By focusing on both single crop and multi-crop disease detection, the project has demonstrated the effectiveness and versatility of CNN architectures like ResNet and Inception. These models, trained on meticulously collected and preprocessed datasets, have shown high accuracy and reliability in identifying diseases in crops, which can lead to reduced yield losses and improved agricultural productivity.[7] [5]

Moreover, the project's emphasis on real-time crop disease detection further enhances its practical utility. By developing models capable of detecting diseases in real-time, farmers can swiftly respond to disease outbreaks, implementing targeted interventions to mitigate crop losses. This real-time capability is crucial for effective crop management and decision-making, underscoring the project's potential impact on agricultural sustainability and food security.

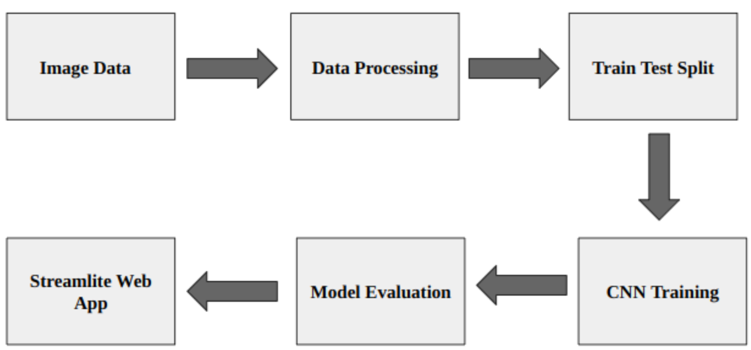

Figure 1: Flowchart of the Methodology used.

4 Dataset

For the single disease detection task, a custom dataset was curated and uploaded to Google Drive. The dataset comprises high-resolution images of Cotton plants and it's leaf affected by disease, as well as images of healthy plants for comparison. The dataset was meticulously annotated to ensure accurate labeling of disease presence and severity levels. Preprocessing steps such as resizing, normalization, and augmentation were applied to standardize the dataset and enhance model training.

For the multi-crop disease detection task, a publicly available dataset from Kaggle was utilized. The dataset contains images of various crops, each with its unique disease characteristics. Similar preprocessing steps were applied to this dataset to maintain consistency and quality across different crop types.

Both datasets were divided into training, validation, and testing sets to facilitate model training and evaluation. Links can be found at ”Assets for Codes, App and Dataset”

5 Algorithms

Algorithm for ResNet-152V2 Model for Single Crop Disease Detection[4] is explained below:

Input Layer: The input layer is a convolutional layer with a kernel size of 7x7 and a stride of 2, followed by a max-pooling layer with a stride of 2.

Convolutional Blocks: The network consists of multiple convolutional blocks, each containing two convolutional layers. The first convolutional layer in each block is followed by a batch normalization layer and a ReLU activation function. The second convolutional layer in each block is also followed by a batch normalization layer and a ReLU activation function.

Residual Connections: The output of each convolutional block is added to the input of the block, creating a residual connection. This allows the network to learn more complex representations by combining the outputs of different layers.

Downsampling: The network includes downsampling layers, typically consisting of a convolutional layer with a stride of 2 followed by a max-pooling layer with a stride of 2. These layers reduce the spatial dimensions of the feature maps.

Output Layer: The final output layer is a global average pooling layer followed by a fully connected layer with a softmax activation function for classification tasks.

Algorithm for Building Multiple Crop Disease Detection Model using CNN and Keras[7] [5] workflow is explained in detail below:

Set Seeds for Reproducibility: Import necessary libraries like random, numpy, and tensorflow. Set random seeds for random, numpy, and tensorflow to ensure consistent results across runs.

Import Dependencies: Import additional required libraries and modules such as os, json, ZipFile, PIL, matplotlib, and tensorflow.keras.

Data Preprocessing: Define the base directory for the dataset as 'plantvillage dataset/color'. Load and display an example image from the dataset using matplotlib.image.imread() and matplotlib.pyplot.imshow(). Observe the shape of the image, which is (256, 256, 3).

Model Building: Define the CNN model architecture using TensorFlow's Keras API. Create the model by stacking convolutional, pooling, and dense layers. Compile the model with an appropriate loss function, optimizer, and evaluation metrics.

Model Training: Preprocess the images using ImageDataGenerator for training and validation. Split the dataset into training and validation sets. Train the CNN model on the training data using the fit() method. Monitor the training and validation accuracy and loss during the training process.

Model Evaluation: Evaluate the trained model's performance on the validation set using the evaluate() method. Visualize the training and validation accuracy and loss curves using matplotlib.pyplot.

Model Testing: Test the trained model on unseen data to predict plant diseases. Save the trained model for future uses.

6 Results

The ResNet-152V2 model, employed for single crop disease detection, demonstrated a commendable accuracy in finding the difference between healthy and diseased cotton plants. The model's accuracy and recall were notably high, showing us its efficiency in disease identification. On the other hand with the multi-crop disease detection, using the Keras model exhibited performance across various crops. It showed slightly lower accuracy than the ResNet model, despite that it can be trusted and implemented because of its versatility. The real-time disease detection app, developed as part of the project, underwent iterative refinement processes to improve its accuracy, achieving an impressive accuracy of over 90% in identifying plant diseases. The user-friendly interface and real-time disease alerts of the app enhance its utility for farmers in monitoring and managing crop health.

Below are the snapshots of various results and model performance graphs.

Single Crop Disease Detection:

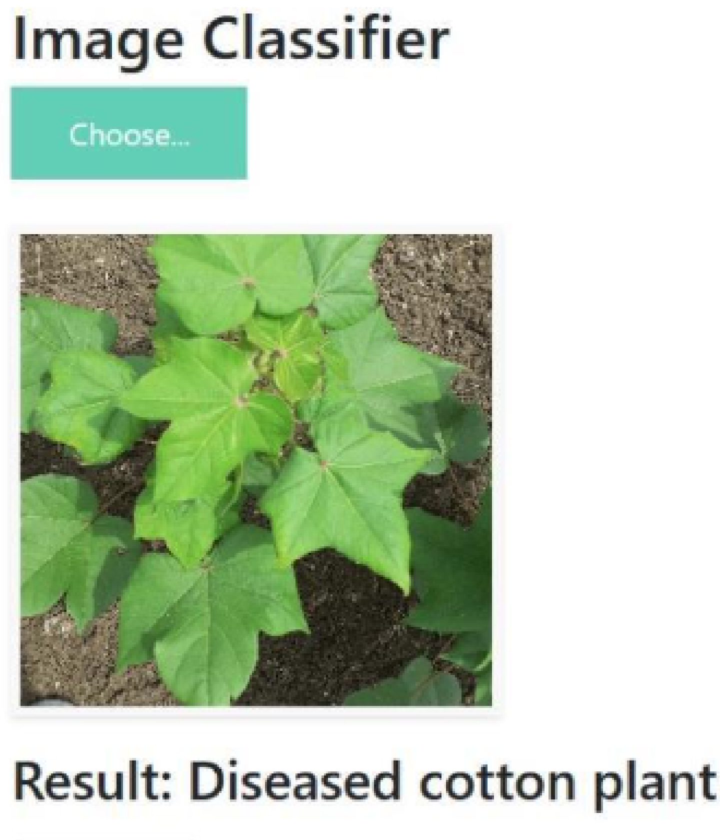

Figure 2: Output for disease ridden Cotton Plant

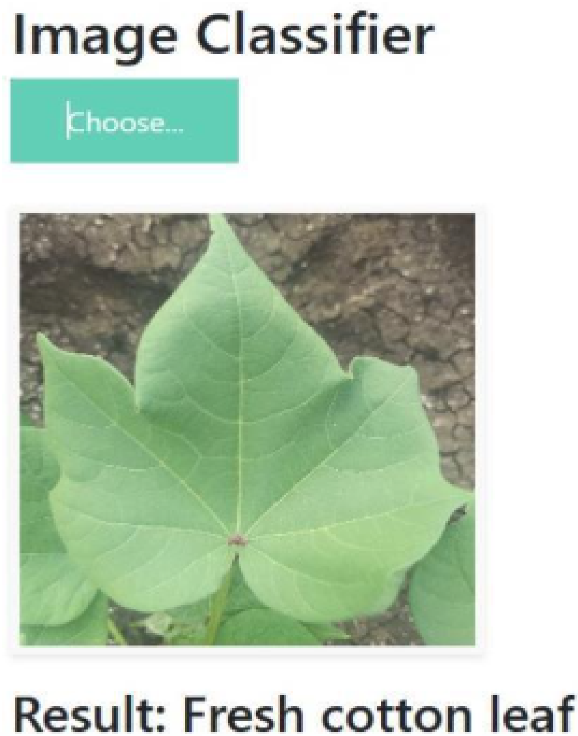

Figure 3: Output for Healthy Cotton Plant

Figure 4: more results for Cotton Plant.

Multi-Crop Disease Detection:

Figure 5: for Potato crop.

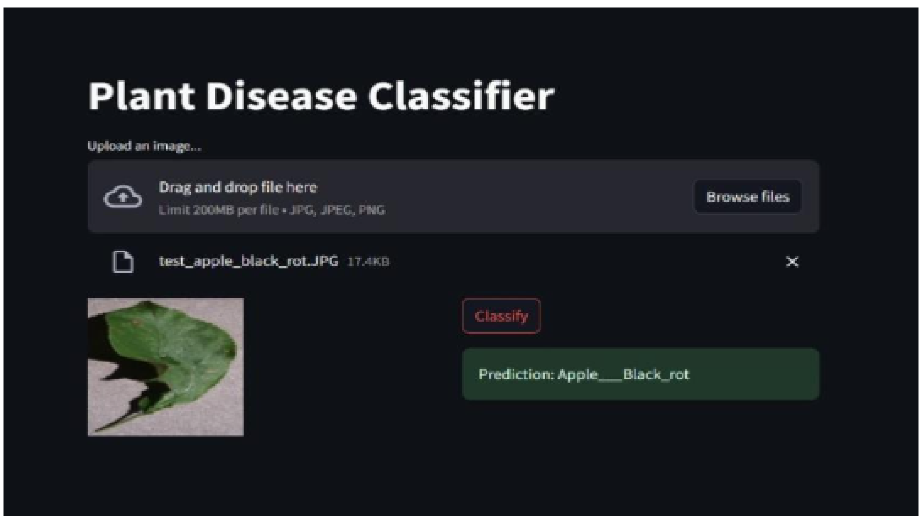

Figure 6: for Apple crop.

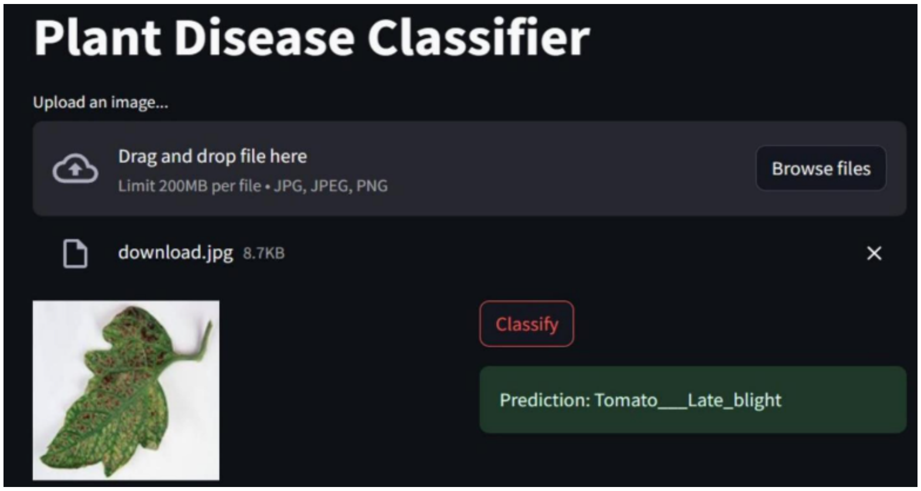

Figure 7: for Tomato crop.

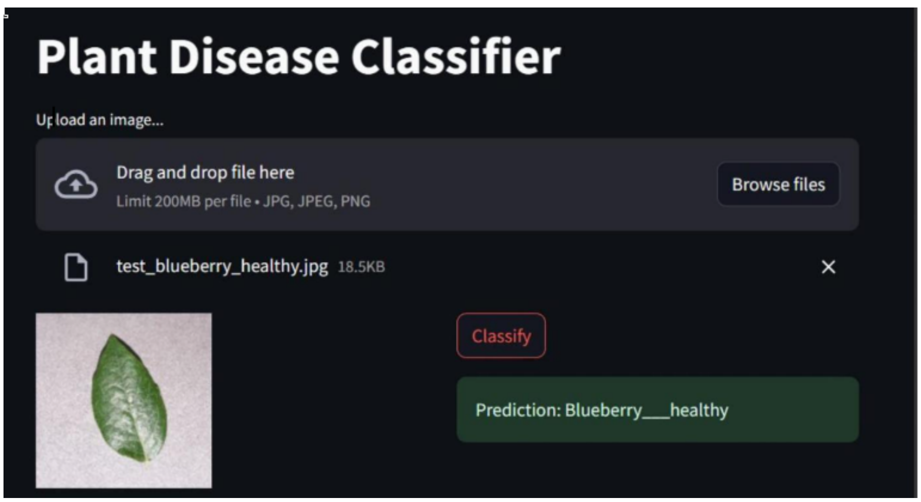

Figure 8: for Blueberry crop.

Model Performance Graphs:

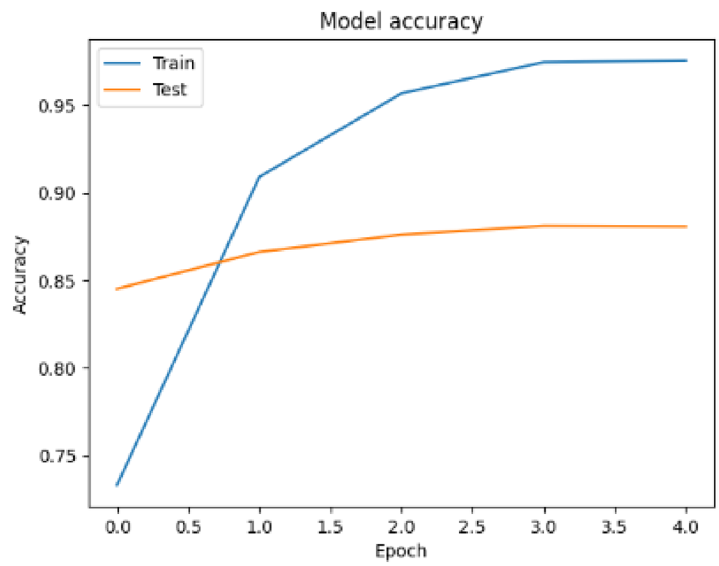

Figure 9: Graph showcasing trend of Model Accuracy

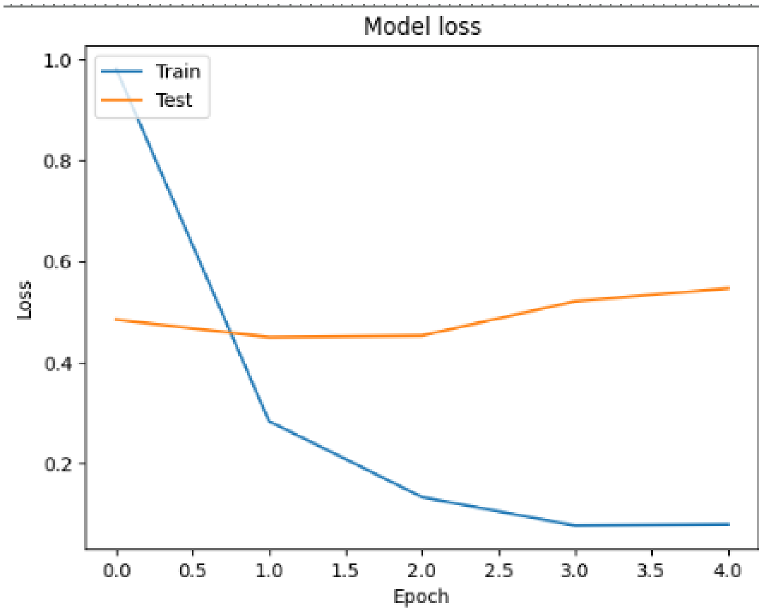

Figure 10: Graph showcasing trend of Model Loss

7 Discussion

In collecting a dataset for disease detection in crops, several challenges arise. The first involves identifying the target crops, as different crops exhibit varying disease characteristics, necessitating a diverse dataset for robust model training. Next is the selection of disease types, as each crop can be susceptible to multiple diseases, requiring a comprehensive dataset. Additionally, ensuring an adequate dataset size, with a sufficient number of images for each class (healthy and diseased), is crucial but resource-intensive. Data quality is another challenge, as ensuring consistency through preprocessing is essential yet challenging when dealing with images from different sources. Including diverse data from various geographical locations and growth stages is vital for model generalization but challenging to obtain. Lastly, efficient data management, including organization and security, is necessary, particularly with a large dataset and limited resources.

In training a model for crop disease detection, challenges arise due to data scarcity and variability. Training Convolutional Neural Networks (CNNs) requires large, diverse datasets with images representing various crop types, disease conditions, and environmental factors, which can be expensive and time-consuming to collect and annotate. Potential solutions include collaborating with agricultural institutions or organizations to access existing datasets, using data augmentation techniques to expand the dataset, and exploring transfer learning with pre-trained models. Additionally, variability in disease symptoms and crop appearances due to factors like lighting conditions and growth stages can lead to poor model performance on unseen data. To address this, collecting diverse datasets, implementing robust preprocessing techniques, and using data augmentation to simulate different conditions can improve model generalization and robustness.

Problems were faced in the deployment of a real-time disease detection app included low accuracy (77%) initially, which was improved by strategies such as increasing the size and diversity of the training dataset, fine-tuning model hyperparameters, and utilizing advanced CNN architectures or transfer learning techniques. Hosting and server connectivity issues also arose, requiring proper configuration of server and networking settings and implementation of error handling mechanisms. To address these challenges, a Streamlit web app for disease detection was developed, achieving an accuracy of over 90% for different disease detection tasks. This success was attributed to further model refinement based on feedback and iterative testing, as well as user-friendly interface design enhancing user experience and application usage.

8 Conclusion

The project utilizes Convolutional Neural Networks (CNNs) for crop disease detection represents a valid approach to modern agriculture. By harnessing the power of deep learning, the project aims to address the challenge of timely and accurate identification of crop diseases faced by many farmers worldwide. The capability is crucial for maintaining crop health and maximizing yields, ensuring food security for the growing world population at whole.

In the **Single Detection phase**, the project demonstrates the effectiveness of the ResNet-152V2 model at classifying diseases in cotton plants, accurately. Model's architecture having deep layers and subsequent residual connections, allowed it to learn intricate patterns and features from images, enabling it to differentiate between healthy and diseased plants with high accuracy. Our implementation of this model showcases its potential as a tool for early disease detection and targeted treatment, Successfully.

In the **Multi Detection phase**, the project extends its focus to a wider range of crops, highlighting the versatility of the Keras model. While not as specialized as the ResNet-152V2 model, it demonstrates comparable performance across different crop types. Its flexibility and ease of implementation make it a practical choice for farmers dealing with diverse crop diseases, offering a more generalized yet effective solution for disease detection.

The project's integration of real-time monitoring and IoT devices further enhances its utility in agricultural settings. By providing farmers with instant feedback on crop health and disease status, the project enables decision-making and timely interventions. This capability is invaluable in mitigating the spread of diseases and minimizing crop losses, ultimately leading to improved yields and economic benefits for farmers. The project's future scopes, including advanced flight planning and agriculture integration, hold immense potential for further optimizing crop management practices. By exploring these avenues, the project can enhance its capabilities in terms of efficiency, scalability, and sustainability, making it a valuable asset for farmers and agricultural communities worldwide.

In conclusion, the project represents a significant step forward in the field of precision agriculture, showcasing the potential of advanced technologies like CNNs, IoT, and drone technology to transform agricultural practices. By continuing to innovate and collaborate, the project can pave the way for a more resilient and productive agricultural sector, ensuring food security and environmental sustainability for generations to come.

9 Future Scope

These future scopes can further enhance the capabilities of your project, making it more effective and scalable in addressing challenges in agriculture and drone technology integration.

Drone Technology Integration: Advanced Flight Planning involves developing algorithms for optimized flight paths to maximize coverage and minimize energy consumption. Swarm Intelligence explores the use of multiple drones working collaboratively to cover larger areas efficiently. Autonomous Navigation implements machine learning algorithms for obstacle avoidance and safe navigation in complex environments.[9]

Real-time Monitoring: Automated Disease Recognition enhances machine learning models to detect and classify diseases in real-time, enabling immediate action. Alert Systems are implemented to trigger automated alerts based on disease severity levels, enabling timely interventions. Integration with Agricultural Machinery involves integrating drone data with existing machinery for automated responses, such as pesticide application.

Agriculture Integration: Variable Rate Application utilizes drone data to implement variable rate application of fertilizers and pesticides, optimizing resource use. Yield Mapping integrates drone data with yield mapping technologies to correlate crop health with productivity for precise management.[2][4]

Data fusion with IoT Systems: Multispectral Image Fusion develops algorithms to fuse multispectral images from drones with other sensor data for comprehensive analysis. IoT Sensor Integration integrates IoT sensors for soil moisture, weather conditions, and crop growth stages to enhance decision-making.

Collaborative Research: Partnerships with Agricultural Institutions involve collaborating with agricultural research institutions for data sharing and model validation. Industry Partnerships entail partnering with drone technology companies for access to cutting-edge hardware and software solutions. Providing leeways to food security.

10 References

[1] Afroza Akter and A K M Sikder. "A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends". In: *Sensors* 23.10 (2023), p. 4769.

[2] Lavanya Ganesan and Sushma Pandey. "Smart agriculture with machine learning and deep learning: A review". In: *Agricultural Reviews* 39.4 (2020), pp. 387–413.

[3] Manish Jha, Akshay Prajapati, and Pravin Kumar Arya. "Plant disease detection and classification techniques: a comparative study of the performances". In: *Journal of Big Data* 6.1 (2023), pp. 1–18.

[4] Andreas Kamilaris, Xavier Prenafeta-Lloret, and Jordi Jordà-Sellers. "Deep learning for image recognition in precision agriculture". In: *Frontiers in Plant Science* 9 (2018), p. 1719.

[5] Geert Litjens et al. "Computer-aided detection of prostate cancer in MRI scans: a review of the literature". In: *Computerized Medical Imaging and Graphics* 38.7 (2014), pp. 789–802.

[6] Yang Lu, Shaowei Wu, and Yiran Wang. "The challenges of image classification for autonomous agricultural robots". In: *Computers and Electronics in Agriculture* 152 (2018), pp. 78–88.

[7] Sharada P Mohanty, David P Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection". In: *Frontiers in Plant Science* 7 (2016), p. 1414.

[8] Imen Nadal et al. "Deep learning for plant disease detection: A survey". In: *Computers and Electronics in Agriculture* 178 (2020), p. 105787.

[9] Jennifer M Pound et al. "Deep learning for plant disease detection and classification using UAV-acquired multispectral imagery". In: *The Plant Journal* 96.1 (2018), pp. 187–202.

[10] Connor Shorten and Taghi M Khoshgoftaar. "A survey on image data augmentation for deep learning". In: *Journal of Big Data* 6.1 (2019), p. 60.

[11] Guanqi Wang et al. "Deep learning for smart agriculture: a survey". In: *Agricultural Sciences* 12.07 (2021), pp. 818–831.

11 Assets for Codes, App and Dataset

Single Crop Disease Detection: `https://github.com/CodeWizard812/Cotton-Disease-Detection`

Multi Crop Disease Detection: `https://github.com/CodeWizard812/Multiple-Crop-Disease-Detection`

Real-time Disease Detection App: `https://github.com/CodeWizard812/Real-Time-Plant-Disease-Detection`

Dataset for Single Disease Detection: `https://drive.google.com/file/d/1u1CxT1EvxaaGKacho55kHWJkEe0rI1Xw/view?usp=drive_link`

Dataset for Multiple Disease Detection: `https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset`