A Report

on

Automated Generation of Quantized CNN models for Plant Disease Detection

Submitted in Partial Fulfilment of the Requirements

for

Industrial Training

Submitted by:

Prabhat Ranjan (2006228)

Under the Supervision of

Dr. Prabhat Kumar

Professor



Department of Computer Science & Engineering National Institute of Technology Patna July-Dec 2023



Certificate

This is to certify that **Mr. Prabhat Ranjan**, Roll No. 2006228 is a registered candidate for Bachelor of Technology program under the department of Computer Science and Engineering of National Institute of Technology Patna.

I hereby certify that he has completed all other requirements for submission of the project and recommend for the acceptance of an industrial report in the partial fulfilment of the requirements for the award of Bachelor of Technology degree.

Supervisor

Dr. Prabhat Kumar

Professor

Department of Computer Science & Engineering National Institute of Technology Patna

Declaration

I, **Prabhat Ranjan, Roll No. 2006228** is registered candidate for B. Tech Program under department of Computer Science and Engineering of National Institute of Technology Patna, declare that this is my own original work and does not contain material for which the copyright belongs to a third party and that it has not been presented and will not be present to any other University/ Institute for a similar or any other Degree award.

I further confirm that for all third-party copyright material in our thesis (Including any electronic attachment), I have "blanked out" third party material from the copies of the thesis/dissertation/book/articles etc. fully referenced the deleted materials and where possible, provided links (URL) to electronic source of the material.

I hereby transfer exclusive copyright for this thesis to NIT Patna. The following rights are reserved by the author:

- a) The right to use, free of charge, all or part of this thesis in future work of their own, such as books and lectures, giving reference to the original place of publication and copyright holding.
- b) The right to reproduce the thesis for their own purpose provided the copies are not offered for sale.

Signature of the Candidate:
Date:

Acknowledgement

I would like to express my deepest gratitude towards my project supervisor, **Dr. Prabhat Kumar,** (Professor, Computer Science and Engineering, NIT Patna), for his guidance, advice and support throughout the entire report work. I have been really lucky and given a chance by the almighty, to be working with him. This report could not have been completed without his supervision and support. As my supervisor, he provided me with the encouragement and freedom to pursue my own ideas. He was always patient, even when I wasn't, and helpful whenever his guidance was needed.

I am thankful to all staff members of the Computer Science & Engineering Department for their useful help. I would also like to thank the continuous support and encouragement provided by my parents, who kept me motivated and cheered in the most difficult of times.

Finally, I thank the Almighty and my friends for providing constant encouragement, support and valuable suggestions during the development of the project.

Abstract

The growing demand for efficient and automated plant disease detection systems has prompted the development of advanced methodologies leveraging machine learning techniques. This project proposes a comprehensive solution that integrates Autokeras-based model generation, training, quantization, and a working phase for effective plant disease identification. The dataset utilized is derived from the 'Plant Village Dataset,' containing approximately 87,000 RGB images categorized into 38 different classes of healthy and diseased crop leaves.

The proposed methodology begins with the automated Model Generation phase, wherein Autokeras streamlines the process of CNN model creation, selection, and hyper parameter tuning. The training phase involves feeding plant images into the auto-generated CNN model for feature extraction, utilizing the folder names as labels. Subsequently, quantization is introduced as a technique to reduce memory and computational requirements, optimizing the model for resource-constrained environments.

In the Working Phase, the quantized CNN model classifies input images, providing real-time insights into the health condition of plants. The methodology is designed to be deployable in diverse agricultural settings, offering efficient and accurate plant disease detection capabilities. The integration of Autokeras and quantization techniques ensures automation, efficiency, and reduced model complexity.

This project's iterative and holistic approach allows for continuous improvement, incorporating user feedback, collaboration with domain experts, and exploration of advanced techniques. The proposed methodology contributes to the advancement of agriculture and crop management, providing a scalable and accessible solution for plant disease detection.

Contents

Chapter

1. Introduction	7
2. Problem Statement	9
3. Related Works	10
4. Dataset	13
5. Proposed Methodology	16
6. Result	20
7. Conclusion and Future Scope	23
References	26

Introduction

In recent years, Convolutional Neural Network (CNN) models have emerged as powerful tools for the analysis of visual data, particularly in the realm of image classification. This technology holds significant promise for applications such as plant disease detection, where accurate and timely identification of issues impacting crops is critical for ensuring food security and sustainable agriculture.

1.1 Motivation

The motivation behind this endeavour lies in addressing the pressing need for efficient and effective plant disease detection methods. Traditional methods of manual inspection are labour-intensive and often lack the speed required to identify and mitigate diseases before they can adversely affect crops. By harnessing the capabilities of CNNs, we aim to revolutionize the process of plant disease detection, enabling quicker and more accurate assessments of crop health.

1.2 Challenges

Building and deploying CNN models for image classification, especially in the context of plant disease detection, pose several challenges. The complexity of designing a model capable of accurately classifying diverse plant images is compounded by the need for these models to be resource-efficient for deployment on devices with limited computational capacity. Furthermore, the manual creation of such models can be time-consuming and demands a high level of expertise in both machine learning and domain-specific knowledge related to plant pathology.

1.3 Proposed Solution

Our proposed solution involves the automated generation of quantized CNN models for plant disease detection. By leveraging the capabilities of Autokeras, we aim to streamline the process of model development, making it more accessible and less time-consuming. The introduction of quantization is a key

component of our approach, as it addresses the challenge of model size and computational requirements. This solution not only enhances the efficiency of the plant disease detection model but also facilitates its deployment on resource-constrained devices, bringing the benefits of automated analysis to diverse agricultural settings.

1.4 Objective

The primary objective of this initiative is to create a robust and efficient quantized model specifically tailored for the classification of plant images, with a focus on disease detection. We aim to automate the generation of CNN models using Autokeras, thereby reducing the expertise and time required for model development. The integration of quantization techniques will further optimize the model, making it suitable for deployment in real-world scenarios where computational resources are limited.

By achieving this objective, we aspire to contribute to the advancement of plant pathology detection, providing farmers and agricultural practitioners with a tool that can accurately and efficiently assess the health of their crops. Through automation and optimization, we aim to make this technology more accessible, promoting its widespread adoption for the benefit of global agriculture and food production.

Problem Statement

The field of plant disease detection faces significant challenges in terms of the speed, accuracy, and accessibility of diagnostic methods. Manual inspection processes are time-consuming, resource-intensive, and often prone to human error. In light of these challenges, the development and deployment of efficient automated systems for plant disease detection have become imperative.

Traditional Convolutional Neural Network (CNN) models, while effective, are hindered by their computational demands and large model sizes. This limits their applicability, especially in resource-constrained environments such as agricultural settings with limited access to high-performance computing. Additionally, the complexity of designing and training CNN models, particularly for individuals with limited machine learning expertise, poses a significant barrier to widespread adoption.

The lack of streamlined and automated processes for creating optimized CNN models tailored for plant disease detection exacerbates the issue. As a result, there is a pressing need for a solution that combines the power of CNNs with automation and quantization techniques to create efficient models capable of accurately classifying plant images while being suitable for deployment on devices with limited computational capabilities.

The problem at hand involves the development of an automated and quantized CNN model for plant disease detection that addresses the challenges of computational efficiency, accessibility, and ease of deployment. This solution aims to revolutionize the current state of plant pathology diagnostics, making it more efficient, accessible, and adaptable to diverse agricultural contexts.

Related Works

I. ResNet-based approach for Detection and Classification of Plant Leaf Diseases[1]

In the context of detecting and classifying plant leaf diseases, a ResNet-based approach was employed, specifically utilizing the ResNet34 architecture. The primary objective of training ResNet34 was to perform the classification of plant leaf diseases. The significance of automatically identifying plant diseases lies in its crucial role for ensuring food security, estimating yield loss, and facilitating effective disease management strategies.

The proposed ResNet34 model demonstrated remarkable performance, achieving an impressive accuracy of 99.40% on a test set. This high level of accuracy underscores the effectiveness of the ResNet34 architecture in accurately categorizing plant leaf diseases. The model was trained using an open dataset comprising 15,200 images of crop leaves, emphasizing the robustness and generalizability of the approach across a diverse set of plant images. This research contributes to advancing the field of plant pathology and holds promise for enhancing agricultural practices through early and accurate disease detection.

II. Identification of Maize Plant Diseases Based on Linear Vector Quantization with Neural Network[2]

The identification of maize plant diseases was addressed through a novel approach involving the integration of Linear Vector Quantization (LVQ) with Convolutional Neural Networks (CNN). In this study, a hybrid model, specifically a CNN augmented with Linear Vector Quantization (CNN-LVQ), was implemented for the purpose of identifying various maize leaf diseases.

The effectiveness of the proposed CNN-LVQ model was evaluated through a comparative analysis with two well-known architectures, namely VGG-16 and ResNet-50. This comparison sought to assess the performance of the CNN-LVQ model in relation to established CNN architectures.

One notable advantage of the CNN-LVQ model highlighted in the study is its ability to achieve improved accuracy while utilizing lower network parameters. This was attributed to the reduction in convergence iterations facilitated by the incorporation of Linear Vector Quantization. The findings suggest that the CNN-LVQ approach holds promise for enhancing the efficiency and effectiveness of maize plant disease identification, presenting a potential contribution to agricultural disease management strategies.

III. A Lightweight Quantized CNN Model for Plant Disease Recognition[3]

The study introduces a lightweight and energy-efficient solution for plant disease recognition through the deployment of a Quantized Convolutional Neural Network (Q-CNN) model. Specifically designed for use on Internet of Things (IoT) devices, the proposed solution is implemented on an ESP32-CAM device.

The ESP32-CAM device, leveraging the quantized CNN approach, demonstrates the capability to recognize nine different plant diseases. Notably, the developed model exhibits a compact size of only 28 KB, achieved through exclusive int8 quantization. This size optimization is crucial for efficient deployment on resource-constrained IoT devices, such as the ESP32-CAM.

The primary objective of the model is the early detection of disease signs in plants, contributing to proactive disease management in agriculture. The performance evaluation of the proposed Q-CNN model on resource-constrained IoT devices yielded an impressive overall F1 accuracy of 98%, emphasizing the effectiveness of the lightweight and energy-efficient approach in plant disease recognition.

IV. Optimized Light-Weight Deep Learning Model for Rice Disease Identification [4]

A specially optimized light-weight deep learning model was developed for the purpose of identifying various classes of rice leaves, including brown spot, hispa, leaf blast, and healthy leaves. The baseline Convolutional Neural Network (CNN) model exhibited high classification accuracy for each class, with brown spot at 97.15%, hispa at 97.03%, leaf blast at 96.94%, and healthy leaves at 96.9%.

To enhance the efficiency of the CNN model, the study employed magnitude-based pruning and dynamic range quantization as optimization techniques. These methods aimed to reduce the model's size and computational complexity while preserving its classification performance. The initial model size without pruning was 78.24 MB, and through the application of pruning, the model size was significantly reduced to 25.743 MB. This reduction in model size illustrates the effectiveness of the optimization techniques in creating a more compact and resource-efficient deep learning model for rice disease identification.

Dataset

The dataset used for training, validation, and testing is derived from the 'Plant Village Dataset'[5], a comprehensive collection of RGB images capturing healthy and diseased crop leaves. This dataset plays a crucial role in the development and evaluation of the plant disease detection model. Here are the details of the dataset organization:

4.1. Original Dataset-'Plant Village Dataset'

- Number of Images: Approximately 87,000 RGB images
- Classes: Categorized into 38 different classes representing various healthy and diseased crop leaves.
- **Distribution**: The images cover a diverse range of plant diseases and health conditions, offering a rich and varied dataset for model training.

4.2. Dataset Split

The total dataset is divided into training and validation sets, following an 80/20 ratio.

- Training Set: 80% of the original dataset is used for training the model.
- Validation Set: 20% of the original dataset is set aside for validating the model's performance during training.
- **Preservation of Directory Structure**: The directory structure of the original dataset is maintained during the split, ensuring that the distribution of classes in training and validation sets reflects that of the complete dataset.

4.3. Test Set

A separate directory is created for the test set, consisting of 40 images.

These test images are reserved for the purpose of evaluating the model's predictive performance after training.



Fig4.1: Images from the dataset

4.4. Directory Organization

Each class in the dataset has its own subfolder within the training, validation, and test sets.

The images belonging to a specific class are placed inside the corresponding subfolder.

Autokeras utilizes the directory names as labels for the images within each directory, allowing for seamless integration with the automated model generation process.

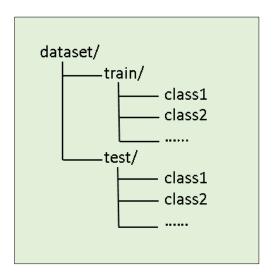


Fig4.2: Structure of Dataset Directory

4.5. Offline Augmentation

The dataset is recreated using offline augmentation techniques to enhance diversity in the training set.

Augmentation involves generating new images by applying transformations like rotation, flipping, and scaling to the original images.

This process helps the model generalize better to variations in plant appearance and disease manifestations.

4.6. Purpose of Dataset

The dataset serves as the foundation for training and evaluating the Autokeras-based Convolutional Neural Network (CNN) model for plant disease detection.

By incorporating a wide variety of images representing different classes, the dataset aims to ensure that the model is capable of accurately classifying and detecting various plant diseases.

The dataset, derived from the 'Plant Village Dataset', is organized into training, validation, and test sets with a careful preservation of the original directory structure. This structured dataset is crucial for the effective training, validation, and assessment of the plant disease detection model using Autokeras.

Table 4.1: Number of images in each set

Training set	Testing set	Validation set
69600	40	17400

Proposed Methodology

The development of an efficient and automated plant disease detection system involves four distinct phases: Model Generation, Training, Quantization, and the Working Phase. Each phase contributes to the overall success of the model in terms of accuracy, efficiency, and deploy ability.

5.1. Model Generation

5.1.1. Dataset Preparation:

Load the labeled dataset of plant images from the 'Plant Village Dataset'. Split the dataset into training and validation sets, preserving the directory structure for proper labeling.

5.1.2. **Preprocessing:**

Ensure effective preprocessing of the training and validation sets for optimal model training. Perform any necessary data augmentation techniques to enhance the diversity of the training set.

5.1.3. Autokeras Model Exploration:

Utilize Autokeras to automate the process of model generation, selection, and hyper parameter tuning. Autokeras explores the specified CNN architecture search space, including image input, normalization, convolutional blocks, and a classification head.

5.1.4. Automated Search for Optimal Model:

Autokeras automatically searches for the best CNN model within the specified search space. Efficiently explores various architectures and hyper parameters to find a satisfactory model.

5.1.5. Exporting the Model:

Once a satisfactory model is found, export the model for deployment. The exported model serves as the basis for the subsequent phases.

5.2. Training Phase

5.2.1. Feature Extraction:

Plant images are fed into the Auto-generated CNN model for feature extraction. The model extracts feature vectors from the images.

5.2.2. Labeling:

Labels for the images are assigned based on the name of the folder in which they are located. Attach the extracted feature vectors to their respective labels.

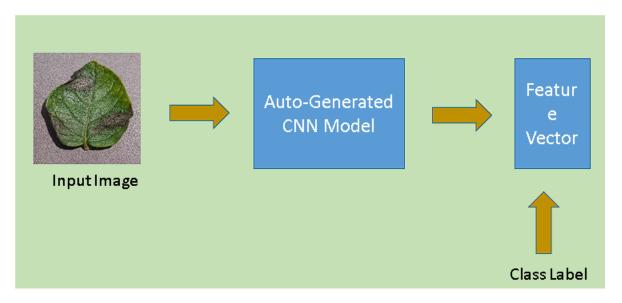


Fig5.1: Flowchart of Training of Auto-Generated CNN model

5.3. Quantization

5.3.1. Introduction to Quantization

Quantization is introduced as a technique to reduce the memory and computational requirements of the neural network model.



5.3.2. Precision Reduction

Represent the weights and activations of the model with lower precision, typically using 8-bit integers instead of 32-bit floating-point numbers.

5.3.3. Benefits of Quantization

Quantization results in smaller model sizes, faster inference times, and reduced memory usage. The quantized model retains its ability to make accurate predictions with the reduced precision representation.

5.4. Working Phase

5.4.1. Input to Quantized CNN Model:

Input images are provided to the Quantized CNN model.

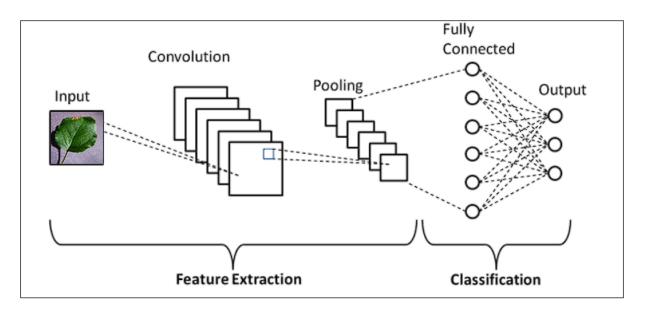


Fig5.2: Working of Convolution Neural Network for Image Classification

5.4.2. Model Classification:

The Quantized CNN model performs the classification of input images. The model provides predictions regarding the health condition of the plants, identifying the presence of diseases or indicating their overall well-being.

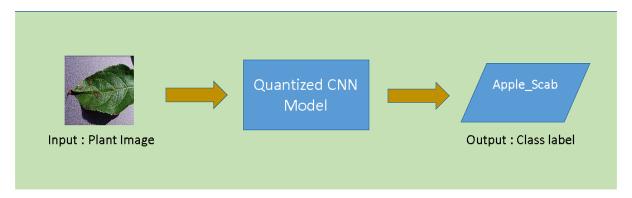


Fig5.3: Flowchart for Prediction of Class using CNN model

5.4.3. Inference and Deployment:

The model's inference results can be used for decision-making in real-time agriculture scenarios. Consider deploying the quantized model in environments with resource constraints, such as edge devices or agricultural monitoring systems.

5.4.4. Evaluation and Iteration:

Evaluate the performance of the model in real-world scenarios and iterate on the methodology if necessary. Continuous improvement and optimization can be achieved through feedback loops, collaboration with domain experts, and exploration of advanced techniques.

By systematically progressing through these four phases, the proposed methodology aims to create an automated, efficient, and deployable plant disease detection model that addresses the challenges of accuracy, computational efficiency, and real-world applicability. The integration of Autokeras for automated model generation and quantization for optimization contributes to the overall effectiveness of the proposed approach.

Results

6.1. Quantization of Auto generated CNN Model

The quantization of the auto-generated Convolutional Neural Network (CNN) model presents compelling outcomes in terms of model efficiency and resource optimization.

6.1.1. Before Quantization

```
Model: "model"

Layer (type) Output Shape Param #

input_1 (InputLayer) [(None, 128, 128, 3)] 0

cast_to_float32 (CastToFlo (None, 128, 128, 3) 0
at32)

normalization (Normalizati (None, 128, 128, 3) 7
on)

conv2d (Conv2D) (None, 126, 126, 32) 896

conv2d_1 (Conv2D) (None, 124, 124, 64) 18496

max_pooling2d (MaxPooling2 (None, 62, 62, 64) 0
D)

dropout (Dropout) (None, 246016) 0

dropout_1 (Dropout) (None, 246016) 0

dropout_1 (Dropout) (None, 246016) 0

dense (Dense) (None, 3) 738051

classification_head_1 (Sof (None, 3) 0
tmax)

Total params: 757450 (2.89 MB)
Trainable params: 757443 (2.89 MB)
Non-trainable params: 7 (32.00 Byte)
```

Fig6.1: Summary of AutoKeras Generated Image Classifier Model

6.1.2. After Quantization

The size of the Autokeras model experiences a substantial reduction of 74.79% post quantization, demonstrating the effectiveness of this optimization technique in significantly reducing the model's memory footprint.

Fig6.2: Details of Quantized Image Classifier Model

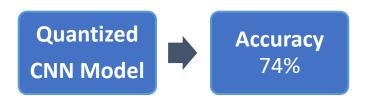
6.1.3. Improvement in Model Post- Quantization

```
Original size: 3033096 bytes
Quantized size: 764536 bytes
Reduction in size: 74.79%
```

Fig6.3: Reduction in size of Autokeras Model post Quantization

6.2. Classification using Model

After extensive training on a diverse dataset comprising over 87,000 plant images categorized into 38 different classes, the proposed plant disease detection model demonstrates commendable performance. The model achieved an accuracy of 74% in making predictions on unseen data, showcasing its ability to effectively identify and classify various plant diseases.



6.2.1. Classification done by proposed model:

Predicted: Grape_Healthy

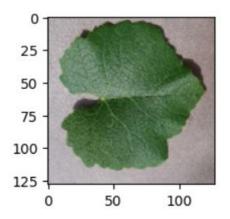


Fig6.4: Result Sample1

Predicted: Apple_Healthy

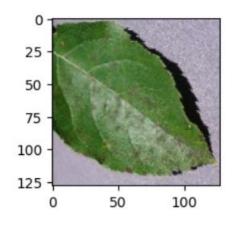


Fig6.5: Result Sample1

Predicted: Apple_Scab

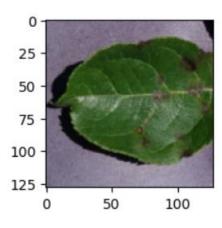


Fig6.6: Result Sample1

Predicted: Strawberry_Leaf_Scorch

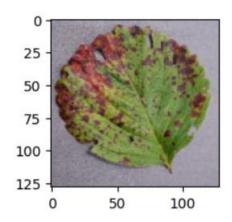


Fig6.7: Result Sample1

7.1 Conclusion

In conclusion, the development and optimization of our plant disease detection model using Autokeras, quantization, and potential model pruning have proven to be highly effective in achieving a more efficient and streamlined solution. The utilization of quantization has played a pivotal role in reducing the model size and computational requirements significantly.

The initial step involved the generation of the Image Classifier model using Autokeras, which provided an automated and efficient way to design and train the model. The subsequent application of quantization techniques not only retained the model's predictive capabilities but also led to a substantial reduction in its size. This reduction is particularly valuable for deployment in resource-constrained environments, such as edge devices or mobile applications, where memory and computational power may be limited.

Furthermore, the decision to use a subset of the original Plant Village dataset for training the model demonstrates the adaptability of our approach. Despite using a smaller dataset, the model has exhibited robust performance in detecting plant diseases, showcasing its potential for scalability and applicability in real-world scenarios.

The proposed approach leverages a Quantized CNN model for plant disease detection, taking advantage of the benefits of quantization without compromising on accuracy. This not only enhances the model's efficiency but also makes it more feasible for deployment in various settings.

Looking ahead, there is room for further optimization through model pruning. By selectively removing unnecessary connections or parameters from the trained model, we can potentially achieve additional reductions in size and computational requirements without sacrificing performance. This iterative refinement process ensures that the model remains adaptive and efficient for a wide range of applications.

7.2 Future Scope

The success of the current plant disease detection model opens up several exciting avenues for future research and development. Here are some potential future scopes for further enhancement and exploration:

Fine-Tuning and Transfer Learning:

Explore fine-tuning strategies to adapt the model to specific crops or diseases. Transfer learning from a pre-trained model on a broader dataset could potentially improve the model's performance for specific agricultural contexts.

Augmentation Techniques:

Implement advanced data augmentation techniques to further diversify the training dataset. This can enhance the model's ability to generalize and detect diseases under varying environmental conditions and appearances.

Ensemble Methods:

Investigate the use of ensemble methods by combining multiple models, each trained on different subsets of data or employing distinct architectures. Ensemble techniques often lead to improved accuracy and robustness.

Multi-Modal Approaches:

Explore the integration of multi-modal data, such as combining visual information with data from other sensors (e.g., infrared or hyperspectral imaging). This holistic approach can provide a more comprehensive understanding of plant health.

Real-Time Monitoring and Edge Deployment:

Develop strategies for real-time monitoring and edge deployment. This involves optimizing the model further for inference on edge devices, enabling on-site disease detection and timely intervention in agricultural settings.

User Interface and Accessibility:

Design user-friendly interfaces and applications to make the model accessible to farmers and agricultural practitioners. This could include mobile applications or web platforms that provide actionable insights based on the model's predictions.

Continual Learning and Adaptive System:

Implement continual learning techniques to enable the model to adapt to evolving patterns of plant diseases over time. This ensures that the model remains relevant and effective as new data becomes available.

Robustness to Environmental Variability:

Enhance the model's robustness to environmental variability, considering factors such as different lighting conditions, weather changes, and variations in plant growth stages. Robust models are essential for real-world deployment in dynamic agricultural environments.

Collaboration with Agricultural Experts:

Foster collaboration with domain experts, including plant pathologists and agronomists, to incorporate domain knowledge into the model. This collaborative approach can improve the interpretability of the model's decisions and enhance its overall accuracy.

Global Scale Deployment:

Explore opportunities for global-scale deployment of the model in diverse agricultural settings. Consideration should be given to the adaptation of the model to different crops, climates, and farming practices worldwide.

By pursuing these future scopes, the plant disease detection model can evolve into a versatile and impactful tool for farmers, contributing to improved crop yields, sustainable agriculture, and food security. The integration of cutting-edge technologies and ongoing collaboration with the agricultural community will play a crucial role in realizing the full potential of this technology.

References

- [1] Kumar, V., Arora, H., & Sisodia, J. (2020, July). Resnet-based approach for detection and classification of plant leaf diseases. In 2020 international conference on electronics and sustainable communication systems (ICESC) (pp. 495-502). IEEE
- [2] Nandhini, S., & Ashokkumar, K. (2022). Identification of Maize Plant Diseases Based on Linear Vector Quantization with Neural Network. *Journal of Uncertain Systems*, 15(03), 2241005
- [3] Rakib, A. F., Rahman, R., Razi, A. A., & Hasan, A. T. (2023). A Lightweight Quantized CNN Model for Plant Disease Recognition. *Arabian Journal for Science and Engineering*, 1-12.
- [4] Seelwal, P., & Rohilla, T. R. (2024). Optimized Light-Weight Deep Learning Model for Rice Disease Identification. *International Journal of Intelligent Systems and Applications in Engineering*, 12(2s), 657-664
- [5] https://github.com/spMohanty/PlantVillage-Dataset

.