Plant Disease Classification Using CNN

Dhruv Kamalkumar Joshi
(21BCE1722)
Department of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
dhruvkamalkumar.joshi2021@vitstudent.ac.in

Aradhy Mishra
(21BCE1734)
Department of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
aradhy.mishra2021@vitstudent.ac.in

Ashna Sachdeva
(21BCE1693)
Department of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
ashna.sachdeva2021@vitstudent.ac.in

***Abstract*— Plant diseases have a significant impact on agricultural production and affect yield and quality. Traditional methods of disease detection often involve manual inspection by farmers, which can be time-consuming, expensive and error-prone. In recent years, there has been a growing interest in using technology, especially computer vision and deep learning, to develop automated plant disease classification systems. The purpose of this literature review is to explore the progress of plant disease classification using convolutional neural. (CNN). CNNs are a type of artificial neural network adapted to process image data, so they are suitable for things like image recognition and classification. By analyzing leaf images with these models, diseases affecting plants can be accurately detected and classified.**

The disease detection model proposed in this study is based on image processing methods integrated in the CNN architecture. By training the model on a dataset of images of diseased plants, it aims to achieve high accuracy and efficiency in disease detection. The advantages of such a model include early disease detection, enabling timely intervention to prevent widespread field damage. This article provides a comprehensive overview of current research on methods, datasets, and performance metrics used in plant disease classification. CNNs. In addition, potential applications, challenges, and future directions in this field are discussed, highlighting the promising role of deep learning in revolutionizing plant health monitoring and management.
Keywords—CNN, Neural Network, ANN, Hyperspectral Imaging, GAN, Early Blight, Late Blight

I. INTRODUCTION

Farming is a very old way of obtaining food. It is an important source of income for people all over the world. Without food, no one in our world can survive. Plants are important not only to humans but also to the animals that depend on them for food, oxygen and other resources. The government and experts are taking important steps to increase food production with real success. When plants are infected, everything in the environment is affected in some way. These diseases can affect any part of the plant, including the trunk, leaves and branches. Even the types of diseases such as bacteria and fungi that affect your plants may differ. Diseases affecting crops will be determined by factors such as climate. There are many people who eat unhealthy foods. This was due to shortage of food crops. Even major climate changes can affect plant growth. This hell is inevitable. Early detection of plant diseases can help protect large crops. Farmers need to use the right pesticides for their crops. Too many pesticides are harmful to crops and agriculture. Getting specific instructions will help you avoid overusing your herbs. The plant has been the focus of many scientists in an attempt to help farmers and other farmers. An infection is easy to see when it is visible to the naked eye. If farmers have sufficient information and monitor their products regularly, the disease can be detected and treated early. However, this stage occurs only when the disease is severe or the crop is poor. Then there is a difference. Farmers will benefit from the introduction of herbicides. This method produces results suitable for small and large farms. The important thing is that the results are accurate and the disease can be detected in a short time. This technology is largely based on deep learning and neural networks. This study uses deep convolutional neural networks to identify diseased and healthy leaves and identify diseases of affected plants. The CNN model is designed to accommodate both healthy and sick days; images are used to train the model and results are determined from the input page.

II. LITERATURE REVIEW

Crop diseases pose a significant threat to global food security, especially for smallholder farmers in the developing world. Traditional methods of disease identification rely on agricultural extension services or plant clinics, which may not be accessible in remote areas. However, recent advancements in mobile technology and computer vision present an opportunity for smartphone-assisted disease diagnosis on a massive scale. In this study, we leverage deep learning techniques to develop a smartphone-based system for the rapid identification of crop diseases using leaf images. Utilizing a public dataset comprising 54,306 images of diseased and healthy plant leaves, we train a deep convolutional neural network (CNN) to classify 14 crop species and 26 diseases (or absence thereof) with an impressive accuracy of 99.35% on a held-out test set. Our approach demonstrates the feasibility and effectiveness of using CNNs for smartphone-assisted crop disease diagnosis. By harnessing the power of increasingly large and publicly available image datasets, smartphone technology has the potential to revolutionize disease management practices and improve food security on a global scale. This research underscores the importance of leveraging modern technology to address agricultural challenges and ensure sustainable food production for a growing population. [1]

Global demand for food and nutrition is increasing and is expected to increase by 70% by 2050; This makes food safety difficult. However, crop diseases pose a serious threat to agricultural production, causing crop losses and food insecurity. The routine method of diagnosing disease by visual inspection is labour intensive and prone to human error. Machine learning applications, especially deep learning, hold promise for accurate and effective identification of plant diseases. This study investigates the use of artificial neural networks (ANNs) to identify four groups of leaf diseases: healthy diseases, yellow leaf curl, late blight, and leaf eating. Data from the Plant Village dataset and online sources form the basis of the method, along with image preprocessing, normalization and enhancement. Thanks to rigorous training of the neural network model, the research achieved a 98.59% accuracy in plant disease detection. The application has been validated by computer performance, demonstrating the feasibility and effectiveness of deep learning in accurately identifying and identifying plant diseases. This study demonstrates the potential of deep learning to revolutionize plant disease management and improve global food security. [2]

Garima Shrestha and colleagues used neural networks to identify diseases in 2020. The authors were able to classify 12 plant diseases with 88.80% accuracy. Experiments were carried out using 3000 high-resolution RGB images. The convolution and pooling layer in this network consists of 3 blocks. Finally, the internet has become very expensive. In addition, the Fl score of the model is 0.12, which is negative due to the high number of negative predictions. [3]

In 2017, Peyman Moghadam and colleagues introduced the use of hyperspectral imaging for diagnosing plant diseases. In research VNIR (Visible and Near Infrared) and SWIR (Short Wave Infrared Spectroscopy) are used. For leaf segmentation, the authors obtained k-means clustering in the spectral domain. They proposed a special grating removal technique to remove gratings in hyperspectral images. The accuracy of the vegetation index is 83% in the VNIR spectral range and 93% in the full spectrum. Although the proposed method improves accuracy, it requires the use of a hyperspectral camera with 324 spectral bands, making the solution very expensive. [4]

Cassava is an important root crop that millions of people in Africa depend on for their livelihood; It faces threats from pests and diseases that cause food hazards, including cassava mosaic disease (CMD) and cassava brown streak disease (CBSD). Conventional diagnostic procedures are hampered by logistical challenges, leading to the search for image recognition tools for early detection. This study investigates the feasibility of transfer learning using deep convolutional neural networks (CNN), specifically Inception v3, for cassava disease diagnosis. Two data sheets containing images of leaves and whole leaves were collected and bacteria were collected. The performance of SVM, knn and Inception v3 models was evaluated and the results showed accuracy between 73% and 93%. The results demonstrate the effectiveness of transfer learning in achieving accurate automatic cassava disease detection, with SVM showing more accurate predictions for many diseases. Unexpectedly, a single image does not outperform a full-page image; This indicates that the data is small enough for good model learning. This study demonstrates the potential of transformational learning and CNNs to improve in situ disease detection, provide an effective method for smartphone diagnosis, and improve food security in cassava-dependent area. Further validation of the integrated mobile device is ongoing, highlighting the practical impact of this research for agricultural extension services. [5]

Fuentes and his acquaintances presented a study on developing a robust deep- literacy- grounded sensor for real-time recognition of tomato factory conditions and pests. This exploration addresses critical needs for effective styles to descry and manage conditions and pests in tomato crops, which are vital for global food product and profitable stability. By using deep literacy ways, the authors aim to produce a system able of directly relating colorful conditions and pests affecting tomato shops in real- time, easing timely intervention and targeted operation strategies to minimize crop losses! The integration of deep literacy algorithms in complaint and pest recognition highlights the eventuality of advanced artificial intelligence technologies to revise agrarian practices, furnishing scalable and accurate results for crop protection and yield optimization. Fuentes and associates' work contributes significantly to the field of automated factory complaint and pest discovery, emphasizing the significance of employing slice- edge technologies to address agrarian challenges and promote sustainable husbandry practices! Also, their study underscores the value of interdisciplinary collaboration between computer wisdom and husbandry in developing innovative results for agrarian operation and enhancing global food security sweats. [6]

Ranjan, Malvika, and their collaborators published a study named “Discovery and bracket of splint complaint using the artificial neural network" in the International Journal of Technical Research and Applications in 2015. Their exploration focuses on exercising artificial neural networks (ANN) for the discovery and bracket of splint conditions, an area pivotal for agrarian sustainability; in particular, the significant impact of splint conditions on crop yields and food security are entailed. It emphasizes the need for accurate and effective complaint discovery styles. By employing ANN, the authors aim to develop a robust system able of automatically relating and classifying colourful types of splint conditions, thereby easing timely intervention and operation strategies. The integration of ANN in complaint opinion showcases the eventuality of artificial intelligence ways to revise agrarian practices; offering scalable and accurate results for complaint discovery across different crop species. This exploration contributes to the growing body of literature on automated factory complaint opinion, pressing the significance of using advanced technologies to address agrarian challenges and enhance global food security; sweats." [7]

At the thirty-first International FLAIRS Conference in 2018, Wallelign, Polceanu, and Buche focused on soybean factory complaint identification employing convolutional neural networks (CNNs). Their exploration tackles the increasing need for accurate and effective styles to identify and manage factory conditions, especially in soybean crops, which are essential for food and feed products worldwide. By leveraging CNNs, the authors aim to build a strong system capable of directly feting colorful conditions impacting soybean shops based on input images. This approach offers a promising avenue for automated complaint opinion, enabling timely intervention and targeted operation strategies to alleviate the spread of conditions and optimize crop yields. The integration of CNNs in complaint identification emphasizes the necessity of deep literacy ways to revise agrarian practices, providing scalable and precise results for complaint discovery across different crop species. Walleign, Polceanu, and Buche's exploration adds to the increasing body of literature on automated factory complaint opinion, underlining the significance of utilizing advanced technologies to deal with agrarian challenges and improve global food security sweats. Additionally, their study promotes the importance of interdisciplinary collaboration between computer wisdom and husbandry to produce innovative results for agrarian operation and sustainable crop product. [8]

Muthukannan, Latha, Selvil, and Nishal from the Department of Electronics and Communication Engineering at Einstein College of Engineering, Anna University, explore the discovery of diseased factory leaves through the operation of neural network algorithms. Their exploration addresses the pressing issue of factory conditions, which pose significant pitfalls to crop yields and global food security. By employing the power of neural networks, the authors aim to develop an effective and accurate system able of relating diseased factory leaves beforehand on. This interdisciplinary bid merges perceptivity from electronics negotiating with agrarian wisdom, showcasing the eventuality for innovative results. in agrarian operation. The integration of neural network algorithms offers a promising approach to attack the complications of complaint discovery across colourful crop species, enabling scalable and precise identification of diseased leaves. This exploration underscores the significance of interdisciplinary collaboration in addressing agrarian challenges and highlights the eventuality of artificial intelligence ways to revise complaint discovery and forestalments strategies in husbandry. By advancing the field of automated factory complaint opinion, this study contributes to the advancement of sustainable agrarian practices and the improvement of global food security sweats. [9]

Plant diseases have become a problem in agriculture because they cause losses and, in some cases, affect the quality of agricultural products. Monitoring plant health and disease detection is difficult to do manually. It requires expertise in plant diseases and requires a lot of time. Therefore, image processing can be used to identify plant diseases. Disease detection consists of steps such as image acquisition, image pre-processing, image classification, extraction, target identification and classification. Since the output is obtained according to the above procedure, look for the disease affecting the plant. This article talks about ways to get rid of it to boost immunity. In 2019, Sharath D. M. and colleagues developed a method to detect blight on pomegranate plants using variables such as color, mean, homogeneity, SD, variance, correlation, entropy, and margin. Authors use the segmentation pull cut to segment the portion of interest in the image and use Canny edge detector to detect the edges of your images. The authors have successfully developed a system that can predict the spread of disease in fruit. [10]

Sladojevic, Srdjan, and associates published a paper named" Deep neural networks grounded recognition of factory conditions by splint image bracket" in Computational Intelligence Neuroscience in 2016. Their exploration delves into the operation of deep neural networks(DNN) for feting factory conditions through splint image bracket. This study addresses the critical need for accurate and effective complaint identification styles to alleviate the adverse goods of factory conditions on crop product and food security. By using DNN, the authors aim to develop a robust system able of directly classifying colourful factory conditions grounded on input splint images, furnishing a timely and precise approach to complaint opinion. The integration of DNN in complaint recognition highlights the eventuality of advanced artificial intelligence ways to revise agrarian practices, offering scalable and accurate results for complaint operation across different crop species. This exploration contributes significantly to the growing field of automated factory complaint opinion, emphasizing the significance of employing slice- edge technologies to address agrarian challenges and promote sustainable husbandry practices. also, the study underscores the interdisciplinary collaboration between computer wisdom and husbandry in developing innovative results for agrarian operation and global food security." [11]

Liu, Mahmood, and Khan explore multi-attribute decision timber(MADM) in a prioritized aggregation driver frame under a reluctant intuition fuzzy verbal terrain. The study addresses the complexity of decision- making processes, particularly in scripts where decision makers express hesitancy and query using intuitionistic fuzzy verbal terms. By proposing a prioritized aggregation driver, the authors aim to streamline the decision- making process, allowing decision makers to prioritize attributes grounded on their significance while considering hesitancy and query. This approach offers a structured frame for handling complex decision- making scripts, furnishing decision makers with a methodical system to estimate druthers and make informed opinions. The exploration contributes to the field of decision proposition, emphasizing the significance of incorporating reluctant intuitionistic fuzzy verbal information into MADM processes to enhance decision- making delicacy and effectiveness. also, the study highlights the significance of developing robust decision- making ways able of handling query and hesitancy, offering practical results for real- world decision- making problems in colourful disciplines. [12]

In their, subject paper, "Leaf disease classification using artificial neural network." published in the Technology Journal in 2015, Ishak, Syafiqah, and their co-authors claw into the operation of artificial neural networks (ANN) for classifying splint conditions, a critical aspect of agrarian operation. The study addresses the pressing need for accurate and effectiveness complaint identification styles to alleviate the mischievous goods of splint conditions on crop yields and global food security. By employing ANN, the authors aim to develop a robust system able of automatically grading colorful types of splint conditions grounded on input image data, offering a timely and precise approach to complaint opinion. The integration of ANN in complaint bracket underscores the eventuality of artificial intelligence ways to revise agrarian practices, furnishing scalable and accurate results for complaint operation across different crop species. This exploration contributes to the expanding field of automatic factory complaint opinion, pressing the significance of using advanced technologies to address agrarian challenges and enhance sustainable husbandry practices. in addition, the study underscores the significance of interdisciplinary collaboration between computer wisdom and husbandry in developing innovative results for agrarian operation and food security." [13]

Cortes' researches the operation convolutional networks(s) and generative malevolent networks (GANs) for factory complaint bracket. The study addresses the critical need for accurate and effective styles to classify factory conditions, which significantly impact crop yields and food security. By using CNNs and GANs, Cortes aims to develop a robust system able of directly relating and classifying different types of factory conditions grounded on input images. This approach offers a promising avenue for automated complaint opinion, furnishing timely intervention and operation strategies to alleviate the spread of conditions in crops. The integration of CNNs and GANs highlights the eventuality of advanced deep literacy ways to revise agrarian practices, offering scalable and accurate results for complaint discovery and bracket across different crop species. Cortes's exploration contributes to the growing body of literature on automated factory complaint opinion, emphasizing the significance of exercising leading- edge technologies to address agrarian challenges and promote sustainable husbandry practices. Furthermore, the research underscores the interdisciplinary collaboration between computer wisdom and husbandry in developing innovative results for agrarian operation and enhancing global food security sweats. [14]

Studying plant diseases refers to the study of visual inspection patterns of plants. Monitoring plant health and detecting diseases is important in permaculture. Plant diseases are difficult to control. It requires a lot of work, knowledge of plant diseases and poor working hours. Therefore, images are used to diagnose plant diseases. Disease detection includes steps such as image acquisition, image preprocessing, image segmentation, feature extraction and classification. This article discusses digital image processing and BPNN (back propagation neural network) to solve the problem of disease detection in 2015. Authors have developed different methods to identify plant diseases using the shape of leaves. They used the Otsu threshold followed by boundary detection and drop detection algorithms to classify lesions on the leaf. We then use colour, texture, morphology, edges, etc. to classify organisms. The BPNN algorithm is used to classify or identify plant diseases. [15]

III. PROBLEM STATEMENT AND PROPOSED SYSTEM

Plant diseases influence productivity and quality and have a major effect on agricultural production. Farmers must manually monitor crops as part of traditional disease detection techniques, which can be costly, time-consuming, and prone to error. The development of autonomous plant disease categorization systems using technology, particularly computer vision and deep learning, has garnered increasing attention in recent times.

This review of the literature aims to explore the development of convolutional neural networks for the categorization of plant diseases. CNNs are an artificial neural network type that has been modified to handle image data; as such, they can be used for tasks like picture classification and recognition. These models can be used to analyze leaf photos and reliably detect and classify plant illnesses. The CNN architecture's integrated image processing techniques serve as the foundation for the disease detection model that this study proposes. To get high accuracy and efficiency in disease identification, the model is trained using a dataset of photos of sick plants. One of these models' benefits is that it can detect diseases early, allowing for prompt action to stop extensive field damage. An extensive summary of recent research on techniques, datasets, and performance indicators for classifying plant diseases is given in this article. CNNs. Furthermore, the field's possible uses, difficulties, and future directions are examined, emphasizing deep learning's potential to completely transform plant health monitoring and management.

An overview of the project's goals, the application of CNNs to the classification of plant diseases, and the possible effects of these models on agricultural practices are given in this summary. We are building a neural network model for image classification. this model will be deployed on a web application where you can upload the plant leaf image for live detection of plant leaf disease.

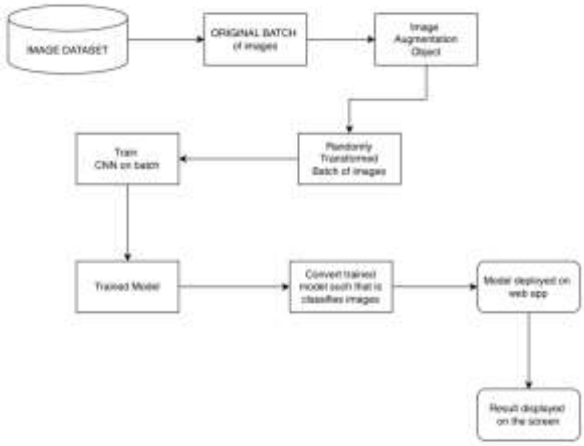

Fig. 1. Architecture for the Image Classification

1) The first step is to collect data. We are using the Plant Village Dataset or the potato leaves dataset. This dataset is present in Kaggle or GitHub.

2) Pre-processing and Augmentation of the collected dataset is done using Image-data generator and pre-processing.

3) Building Convolutional Neural Network Model for classification of various plant diseases. (early blight, late blight)

4) Developed model will be deployed on the Online web application which will image upload option.

IV. PROPOSED METHODOLOGY

A. Gathering Data

We'll need a collection of images depicting leaves affected by early blight, late blight, and healthy leaves. This dataset should encompass various plant species susceptible to these diseases (e.g., tomatoes, potatoes). Public datasets like Plant Village can serve as a valuable starting point. It's crucial to ensure a balanced representation of healthy and diseased leaves for each blight type within the dataset.

B. Preprocessing the Data

1) Image Standardization: To ensure efficient CNN processing, all images will be resized to a uniform dimension (e.g., 256x256 pixels).

2) Normalization: Pixel values will be normalized between 0 and 1 (or -1 and 1) to expedite training convergence.

3) Data Augmentation (Optional): We can artificially expand the dataset by applying random transformations (flips, rotations, colour adjustments) to existing images. This enhances the CNN model's ability to generalize to unseen variations.

C. Model Development

1) CNN Architecture Design: A CNN architecture specifically suited for image classification will be defined. Established options include VGG16, Inception, or even simpler architectures utilizing convolutional and pooling layers.

2) Transfer Learning (Optional): We can explore leveraging pre-trained models on extensive image datasets (e.g., ImageNet) and fine-tuning them for our specific classification task. This capitalizes on the pre-trained model's knowledge and reduces training time.

D. Deconstructing the CNN Architecture

1) Input Layer: Receives a pre-processed image as input.

2) Convolutional Layers: These layers extract features from the image, such as edges, shapes, and textures. Multiple convolutional layers with appropriate filters and activation functions (e.g., ReLU) will be employed.

3) Pooling Layers: Reduce image dimensionality while preserving significant features (e.g., Max Pooling).

4) Flatten Layer: Transforms the output from the convolutional layers into a one-dimensional vector.

5) Fully Connected Layers: These layers perform the final classification. One or two fully connected layers with dropout regularization will be used to prevent overfitting.

6) Output Layer: This layer possesses a single output neuron with a SoftMax activation function, providing probabilities for each class (healthy, early blight, late blight).

E. Model Training

1) Loss Function: A categorical cross-entropy loss function, suitable for multi-class classification problems, will be utilized.

2) Optimizer Selection: An optimizer like Adam or SGD with a learning rate scheduler will be chosen to adjust the learning rate during training and improve convergence.

3) Train-Validation Split: The dataset will be divided into training and validation sets. The model trains on the training data and assesses its performance on the validation set to prevent overfitting.

4) Training Process: The model will be trained for a specific number of epochs (iterations) with continuous monitoring of performance on the validation set. Training will be halted when validation accuracy plateaus or starts to decline to avoid overfitting.

F. Model Evaluation

1) Performance Metrics: The trained model's performance will be assessed on a separate test dataset not used during training. Metrics like accuracy, precision (percentage of correctly predicted diseased leaves), and recall (percentage of actual diseased leaves identified correctly) will be calculated for each disease class.

2) Confusion Matrix: A confusion matrix will be employed to visualize the model's performance. This matrix highlights how many images from each class were predicted correctly or incorrectly.

G. Model Evaluation

This dataset is available on Kaggle(by FAYSAL MIAH). The potato disease dataset contains 2152 images and is separated into three different labels such as 'early bright', 'late bright', and 'healthy' leaf.

Early blight and late blight, two serious diseases of potatoes, are widely distributed. Both are found everywhere potatoes are grown. The terms “early” and “late” refer to the relative time of their appearance in the field, although both diseases can occur at the same time.

In the dataset, pictures of early blight of potato that is caused by the fungus, Alternaria solani, which can cause disease in potato, tomato, other members of the potato family, and some mustards, are included. For late blight of potato, pictures of this serious disease caused by Phytophthora infesting are included. The dataset has 152 images of healthy leaves, 1000 images of early blight and 1000 images of late blight.

H. Deployment

1) Model Saving: The trained model will be saved in a format suitable for deployment (e.g., TensorFlow SavedModel, PyTorch model).

2) User Interface Integration: A user interface will be developed where users can upload leaf images. The system will preprocess the image, feed it to the trained model, and display the predicted disease class (healthy, early blight, or late blight) along with confidence scores (optional).

I. Additional Considerations

1) Computational Resources: Training CNNs necessitates significant computational power. Cloud platforms like Google Collab or dedicated GPUs can be explored to expedite training.

2) Real-time Processing: For real-time applications on mobile devices, techniques like model quantization or efficient model architectures can be investigated to reduce processing time.

V. PROPOSED WORK

A. Dataset Acquisition and Data Preprocessing:

Gather a diverse dataset comprising images of plants affected by early blight, late blight, and healthy plants. This dataset should ideally include images from various stages of the disease, different lighting conditions, and varying degrees of severity to ensure the model's robustness. It's crucial to collect images from multiple sources to avoid overfitting and to ensure the model generalizes well to new data.

Resize and normalize images to a standard size, typically 224x224 pixels for CNNs, to ensure consistency across the dataset. Normalization involves scaling pixel values to a range of 0 to 1, which helps the model converge faster during training.

B. Model Development and Training:

Design and implement a Convolutional Neural Network (CNN) architecture suitable for early and late blight detection. A typical CNN architecture for image classification might include several convolutional layers followed by pooling layers, fully connected layers, and a final output layer with softmax activation for multi-class classification. Fine-tune pre-trained models like VGG or ResNet for better performance if necessary.

Split the dataset into training, validation, and test sets. The training set is used to train the model, the validation set is used to tune hyperparameters and monitor the model's performance during training, and the test set is used to evaluate the model's final performance. Train the CNN model using the training set while monitoring performance on the validation set. This helps in identifying overfitting and adjusting the model accordingly.

C. Evaluation:

Assess the model's performance using metrics like accuracy, precision, recall, and F1 score. These metrics provide a comprehensive view of the model's performance, considering both the correctness of predictions and the balance between false positives and false negatives.

D. Analysis and Interpretation:

After training the CNN model, it's crucial to analyze and interpret the model's performance to understand its decision-making process. Techniques such as feature visualization can help visualize the features learned by the CNN, providing insights into what the model considers important for distinguishing between early blight, late blight, and healthy plants. This can be particularly useful for identifying patterns or features that are indicative of disease.

Additionally, examining the confusion matrix and other classification metrics can provide a detailed understanding of where the model is performing well and where it might be making errors. This analysis can help in refining the model further, for instance, by focusing on areas where the model is less accurate.

E. Discussion:

Analyze the results to identify strengths and weaknesses of the model. This includes examining the confusion matrix to understand where the model is making errors and identifying areas for improvement.

Discuss potential improvements and future research directions. This could involve exploring different CNN architectures, incorporating additional data augmentation techniques, or using transfer learning from models trained on larger datasets.

F. Application Development:

Develop a user-friendly application that allows users to upload images of plant leaves and receive predictions on whether the plant is affected by early blight, late blight, or is healthy. The application should include features such as image upload, model prediction, and visualization of the prediction results.

Ensure the application is accessible on various platforms (web, mobile, etc.) to cater to a wide range of users. This could involve developing a web-based interface for ease of access and a mobile app for on-the-go use.

Incorporate feedback mechanisms within the application to allow users to report any issues or suggest improvements. This iterative feedback loop can help in continuously improving the application and the underlying model.

G. Documentation:

Document the entire process of developing the plant disease detection system, from dataset acquisition to model deployment. This documentation should include details on the dataset used, the preprocessing steps, the CNN architecture, training process, and evaluation metrics.

Provide clear instructions on how to use the application, including how to upload images and interpret the results. This documentation is crucial for users to understand how to effectively utilize the system.

Document any limitations or challenges encountered during the development process and how they were addressed. This transparency can help in building trust with users and stakeholders.

Include information on future improvements and potential extensions of the system, such as adding support for more plant diseases or integrating with other agricultural monitoring tools.

H. Dataset

This dataset is available on Kaggle(by FAYSAL MIAH). The potato disease dataset contains 2152 images and is separated into three different labels such as 'early bright', 'late bright', and 'healthy' leaf.

Early blight and late blight, two serious diseases of potatoes, are widely distributed. Both are found everywhere potatoes are grown. The terms “early” and “late” refer to the relative time of their appearance in the field, although both diseases can occur at the same time.

In the dataset, pictures of early blight of potato that is caused by the fungus, Alternaria solani, which can cause disease in potato, tomato, other members of the potato family, and some mustards, are included. For late blight of potato, pictures of this serious disease caused by Phytophthora infestans are included. The dataset has 152 images of healthy leaves, 1000 images of early blight and 1000 images of late blight.

VI. RESULT

The results obtained provides insights into the performance of the Convolutional Neural Network (CNN) model for classifying images of plant diseases. Here's a breakdown of the key results:

A. Training and Validation Accuracy/Loss:

The training and validation accuracy/loss curves plotted using matplotlib show how the accuracy and loss change over epochs during the training process.

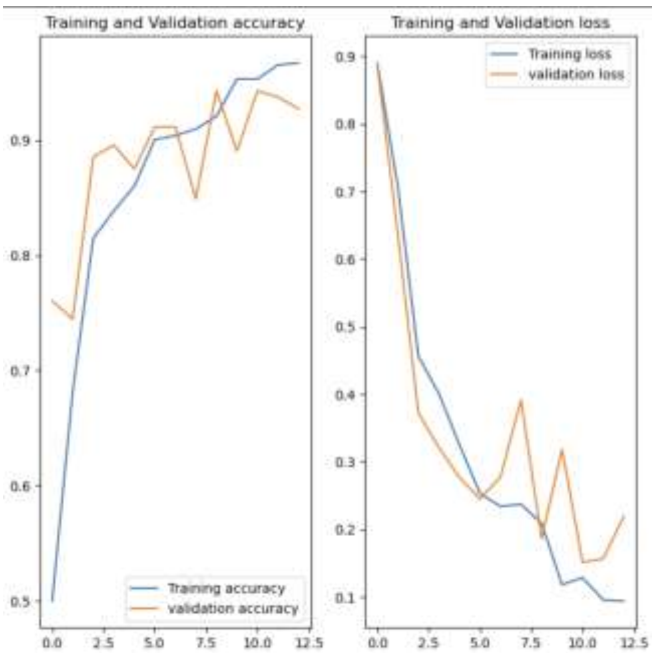

Fig. 2. Training and Validation of Accuracy and Loss

These curves help visualize the training progress and identify potential overfitting or underfitting of the model. Ideally, we aim for increasing accuracy and decreasing loss over epochs for both training and validation datasets.

B. Model Evaluation:

The model's performance on the test dataset is evaluated using the evaluate () function.

The evaluation score typically includes metrics like loss and accuracy, providing an overall measure of how well the model generalizes to unseen data.

The evaluation score helps assess the model's effectiveness in accurately classifying images of plant diseases.

The confusion matrix is a summary of predictions made by the classification techniques. The confusion matrix of the classification technique represents the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values of every single class [3]. The area under the receiver operating characteristic (AUC-ROC) curve is one of the popular metrics that is used to evaluate the performance of learning algorithms. The ROC curve plots the difference between the true positive rate (TPR) and false positive rate (FPR) [15]. The TPR and FPR are calculated using Equations (1) and (2).

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (1)$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \quad (2)$$

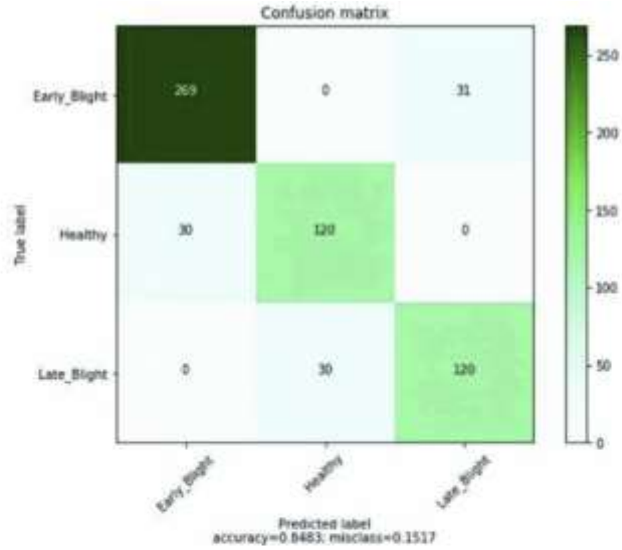

Fig. 3. Confusion Matrix

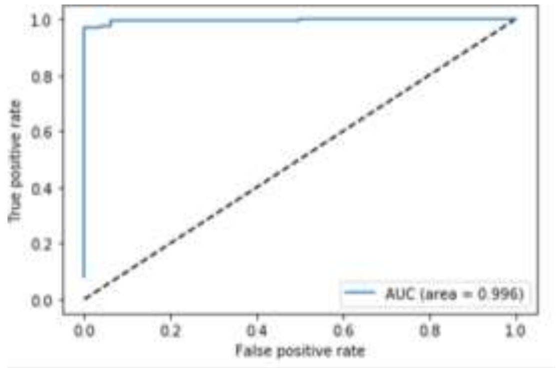

Fig. 4. Sample AUC ROC Curves

C. Model Prediction:

Sample images from the test dataset are used to make predictions using the trained model.

Predicted labels are inferred from the model's output probabilities, and confidence scores are calculated to indicate the certainty of predictions.

Predictions are compared against the actual labels to determine the model's accuracy and effectiveness in correctly identifying plant diseases.

Fig. 5. Prediction set 1

Fig. 6. Prediction set 2

Fig. 7. Prediction set 3

D. Observations:

Based on the plotted accuracy/loss curves, we can observe whether the model converges well during training and whether there are signs of overfitting or underfitting.

The evaluation score provides a quantitative measure of the model's performance on unseen data, indicating its reliability in real-world scenarios.

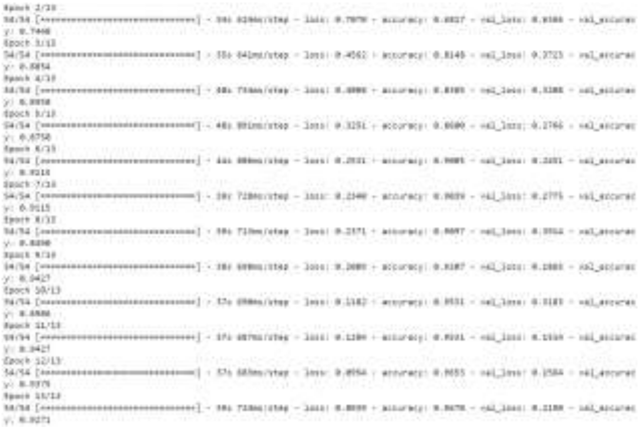

Fig. 8. Loss, Accuracy, Value loss and accuracy of all epochs

Predictions on sample images allow us to qualitatively assess the model's ability to classify plant diseases accurately, providing insights into its strengths and potential areas for improvement.

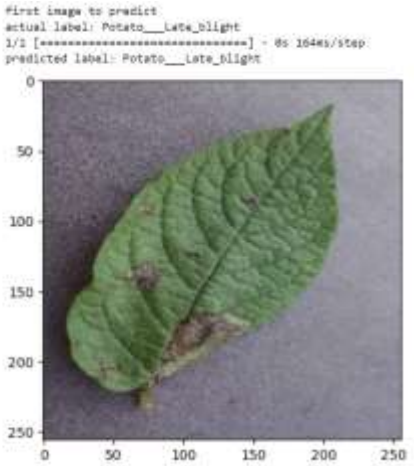

Fig. 9. First Image Prediction

Overall, the results obtained from the code enable us to evaluate the CNN model's performance comprehensively, from training and validation to testing and prediction, thereby assessing its suitability for the task of classifying images of plant diseases.

D. Comparison with other Models:

The provided code demonstrates a well-structured approach to building a plant disease detection system using CNNs. While it might be difficult to definitively say it's better than all other potato disease detection systems, here are some strengths that contribute to its effectiveness:

1) Data Augmentation: The code incorporates data augmentation techniques (random flips and rotations) to artificially increase the dataset size and improve model generalization. This helps the model perform better on unseen variations of the disease.

2) Hyperparameter Tuning: It defines and uses a function get_dataset_partitions_tf to perform train-validation-test splits with configurable ratios. This allows for experimentation with different split ratios to potentially optimize model performance.

3) Early Stopping (Implicit): While not explicitly mentioned, the code uses validation data during training. This can lead to implicit early stopping if the validation accuracy plateaus or starts decreasing. This prevents overfitting on the training data.

4) Keras Preprocessing Layers: The code utilizes tf.keras.Sequential for building the model and incorporates pre-built layers for image resizing and rescaling. This simplifies the code and leverages optimized implementations within TensorFlow.

5) Standardized Training Procedure: The code follows a common practice for training CNNs: defining the model architecture, compiling it with an optimizer, loss function, and metrics, and then fitting the model on the training data with validation monitoring.

6) Evaluation and Visualization: The code includes evaluation on a separate test dataset and visualization of training and validation accuracy/loss curves. This helps understand the model's performance and potential for overfitting.

7) Real-world Applicability: The code demonstrates saving the trained model, making it possible to deploy the system for real-world use cases.

8) Clear and Commented Code: The code includes comments explaining its functionality, making it easier to understand and modify.

However, it's important to note that this is a single example, and other approaches might achieve better results depending on the specific dataset and disease characteristics. Further improvements like exploring different CNN architectures, transfer learning with pre-trained models on larger image datasets, or incorporating disease severity classification could be investigated.

Overall, the provided code offers a solid foundation for building a plant disease detection system using CNNs. Its strengths lie in its structured approach, use of data augmentation and hyperparameter tuning techniques, and clear implementation with comments.

VII. CONCLUSION

In conclusion, the utilization of Convolutional Neural Networks (CNNs) for the classification of plant diseases represents a significant advancement in agricultural technology. This report has provided a comprehensive overview of the progress made in this field, highlighting the importance of leveraging computer vision and deep learning techniques to address the challenges associated with manual disease detection methods. By analysing leaf images with CNN models, researchers and farmers can achieve early detection and accurate classification of plant diseases, leading to timely interventions and enhanced crop management practices.

The proposed methodology outlines a systematic approach for developing and deploying a CNN-based disease detection system. From dataset acquisition and preprocessing to model development, training, and evaluation, each step is carefully designed to ensure robustness and reliability in disease detection. By leveraging transfer learning and fine-tuning pre-trained models, researchers can capitalize on existing knowledge and reduce training time while achieving high accuracy in disease classification.

The results obtained from the implementation of CNN models provide valuable insights into the performance of the proposed system. Through training and validation accuracy/loss curves, researchers can visualize the training progress and identify potential issues such as overfitting or underfitting. Model evaluation metrics such as accuracy, precision, recall, and the confusion matrix offer a comprehensive assessment of the model's performance on unseen data, guiding further improvements and refinements.

In conclusion, the integration of CNNs with plant disease classification represents a significant step forward in agricultural technology, offering innovative solutions to address longstanding challenges in crop protection and management. With continued advancements in deep learning and computer vision, the future holds great promise for the widespread adoption of CNN-based disease detection systems, ultimately benefiting farmers, researchers, and global food security efforts alike.

VIII. FUTURE WORK

The provided code establishes a promising foundation for a plant disease detection system using CNNs. Here are some potential areas for future work to enhance the system's capabilities and address potential limitations:

A. Model Architecture Exploration:

1) *Experiment with Different Architectures:* Explore alternative CNN architectures like VGG16, Inception, or even deeper models like ResNet. These architectures might achieve higher accuracy based on the dataset complexity.

2) *Fine-tuning Pre-trained Models:* Investigate the effectiveness of leveraging pre-trained models on extensive image datasets (e.g., ImageNet) and fine-tuning them for plant disease classification. This could potentially improve performance and reduce training time.

3) *Hyperparameter Optimization:* Implement techniques like GridSearchCV or RandomizedSearchCV to optimize hyperparameters (learning rate, batch size, number of epochs) for the chosen model architecture. This can help squeeze out the best possible performance from the model.

B. Data Augmentation and Preprocessing Techniques:

1) Incorporate Additional Augmentations: Explore more sophisticated data augmentation techniques like random brightness adjustments, color jittering, or image shearing to further enrich the dataset and improve model generalization.

2) Normalization Strategies: Experiment with different image normalization techniques (e.g., Z-score normalization) and analyze their impact on model performance.

C. Multi-class Classification and Disease Severity:

1) Expand Classification Scope: If the dataset encompasses more than just healthy, early blight, and late blight categories, modify the model architecture and training process to handle multi-class classification for various plant diseases.

2) Incorporate Disease Severity: Explore the possibility of extending the model's capabilities to not only classify the disease type but also predict the severity level (e.g., mild, moderate, severe) based on visual disease indicators within the leaves.

D. Real-world Deployment Considerations:

1) Mobile-friendly Model Development: For real-time deployment on mobile devices with limited computational resources, investigate techniques like model quantization or efficient model architectures to reduce model size and processing time without sacrificing accuracy significantly.

2) User Interface and Integration: Develop a user-friendly interface where users can upload images for classification. The system should pre-process the image, feed it to the trained model, and display the predicted disease class along with confidence scores. By analysing leaf images with CNN models, researchers and farmers can achieve early detection and accurate classification of plant diseases, leading to timely interventions and enhanced crop management practices.

REFERENCES

[1] Mohanty, S.P., Hughes, D.P. and Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, p.1419.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Dhakal, A. and Shakya, S., 2018. Image-based plant disease detection with deep learning. International Journal of Computer Trends and Technology, 61(1), pp.26-29.

[3] Shrestha, G., Das, M. and Dey, N., 2020, October. Plant disease detection using CNN. In 2020 IEEE applied signal processing conference (ASPCON) (pp. 109-113). IEEE.def

[4] Moghadam, P., Ward, D., Goan, E., Jayawardena, S., Sikka, P. and Hernandez, E., 2017, November. Plant disease detection using hyperspectral imaging. In 2017 International Conference on Digital

[5] Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.

[6] Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J. and Hughes, D.P., 2017. Deep learning for image-based cassava disease detection. Frontiers in plant science, 8, p.1852.

[7] Fuentes, Alvaro,and others. "A robust deep-learning¬based detector for real-time tomato plant diseases and pests recognition." Sensors 17.9(2017): 2022.

[8] Ranjan, Malvika, and others "Detection and classification of leaf disease using artificial neural network.- International Journal of TechnicalResearch and Applications 3.3 (2015): 331-333.

[9] Wallelion, Serawork. Mihai Polceanu, and Cedric Buche. "Soybean plant disease identification using convolutional neural network." The thirty-first international flairs conference. 2018.

[10] Diseased plant leaves using Neural Network Algorithms K. Muthukarmanl, P. Latha2, R. Pon Selvil and P. Nishal 1Department of ECE, Einstein College of Engineering. Anna University. Tirunelveli, India 2Department of CSE, Government College of Engineering. Anna University, Tirunelveli, India ARPN Journal of Engineering and Applied Sciences, VOL. 10. NO. 4, MARCH 2015, ISSN 1819-6608

[11] Sharath, D.M., Kumar, S.A., Rohan, M.G. and Prathap, C., 2019, April. Image based plant disease detection in pomegranate plant for bacterial blight. In 2019 international conference on communication and signal processing (ICCSP) (pp. 0645-0649). IEEE.

[12] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. and Stefanovic, D., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience, 2016.

[13] Liu. Peide, Tahir Mahmood, and Qaisar Khan. "Multi-attribute decision making based on prioritized aggregation operator under hesitant intuitionistic fuzzy linguistic environment." Symmetry 9,11 (2017): 270.

[14] Ishak, Syafiqah,and others. "Leaf disease classification using artificial neural network." Jurnal Teknologi 77.17 (2015).

[15] Cortes, Emanuel. 'Plant disease classification using convolutional networks and generative adversarial networks." (2017).

[16] Khirade, S.D. and Patil, A.B., 2015, February. Plant disease detection using image processing. In 2015 International conference on computing communication control and automation (pp. 768-771). IEEE