INTERNSHIP PROJECT - II

Project Phase – 3

Image Classification Model and Image Labelling Workflow for Powdery Mildew on Crops Through Machine Learning AI

Submitted To: Dr. Shafaq Khan

Submitted By: Manvir Singh Channa
Harjot Singh Saggu
Guneet Kaur

Introduction

Abstract - **This study presents a novel approach for predicting the severity of powdery mildew infestations on plant leaves using a custom-designed neural network architecture. The model integrates flattened VGG16 features with dense layers and dropout regularization to learn and assess the severity percentage of powdery mildew. Leveraging RGB color information for disease assessment, the architecture exhibits adaptability to disease detection tasks relying on spectral variations. While acknowledging the limitations of RGB-based assessments, the study discusses the potential for incorporating hyperspectral imaging data, highlighting the challenges and opportunities in enhancing disease severity prediction. The output layer, designed explicitly for regression tasks, accurately quantifies disease severity, establishing a robust tool for precise disease management in agriculture. This research contributes to advancing disease assessment methodologies, offering an effective solution to mitigate the economic impact of powdery mildew on crops.**

Keywords – Powdery Mildew, VGG16, Neural Networks, Hyperspectral imaging, Severity Quantification

Background

The timely and accurate identification of powdery mildew severity in plant leaves stands as a pivotal aspect of disease management in agriculture. Powdery mildew, caused by fungal infestations, poses a significant threat to crop yields worldwide. Traditional methods for disease assessment primarily rely on manual visual inspection, prone to subjectivity and inefficiencies in large-scale agricultural settings [1]. To address these challenges, recent advancements in deep learning and image analysis techniques have provided promising avenues for automated disease severity assessment. This study builds upon these advancements, introducing a custom neural network architecture specifically tailored for predicting powdery mildew severity. The proposed architecture integrates VGG16-based feature extraction, leveraging the network's ability to capture intricate visual patterns in plant leaf images. Additionally, this architecture incorporates dense layers with dropout regularization, fostering higher-level feature learning and model generalization. However, while RGB-based assessments offer practicality, the study acknowledges their limitations in capturing spectral nuances inherent in diseases. Thus, this research contemplates the potential integration of hyperspectral imaging data to enrich disease severity prediction. By culminating in an output layer dedicated to regression tasks, the proposed model demonstrates a precision-focused approach to quantifying powdery mildew severity. The development of such a model signifies a significant advancement in disease management strategies, offering a reliable tool for farmers and agricultural practitioners to mitigate the economic impact of powdery mildew infestations. This holistic approach holds the promise of significantly enhancing crop health, yield, and ultimately, the sustainability of agricultural practices.

Problem Statement

Despite advancements in agricultural technology, the timely and accurate assessment of powdery mildew severity remains a challenge in crop management. Manual inspection methods for disease severity evaluation are subjective, labor-intensive, and prone to inconsistencies, leading to suboptimal intervention strategies [1]. This study addresses this critical issue by proposing a novel neural network-based approach that integrates image analysis techniques to predict the severity percentage of powdery mildew infestations on plant leaves. The research aims to provide a robust and automated solution to empower farmers and agricultural practitioners with precise disease severity assessments, fostering improved disease management strategies and enhanced crop health. By leveraging machine learning technology, we aim to address this critical issue. Through advanced algorithm such as CNN and image analysis, we seek to develop a robust model capable of accurately identifying and quantifying the extremity of powdery mildew, providing invaluable insights for targeted intervention strategies.

By addressing these challenges head-on, this research endeavor endeavors to revolutionize greenhouse management practices, ultimately leading to healthier and more productive crops, increased profitability for growers, and a sustainable future for the commercial greenhouse industry.

Research Motivation

The motivation behind this research endeavor stems from the critical need to revolutionize crop disease management, specifically in addressing the challenge of accurately predicting the severity of powdery mildew infestations in plant leaves. Powdery mildew, a devastating fungal disease affecting a wide array of crops, poses a significant threat to agricultural productivity worldwide. The complexity and variability of powdery mildew manifestation, coupled with the labor-intensive and subjective nature of current assessment methods, necessitate innovative and automated solutions for disease severity evaluation.

The inherent challenges in precisely quantifying disease severity hinder effective intervention strategies, leading to suboptimal disease control and substantial economic losses for farmers and agricultural industries. Conventional methods reliant on visual inspection lack precision and scalability, often resulting in delayed or inadequate responses to disease outbreaks [2]. Thus, the development of an automated, accurate, and scalable solution for predicting powdery mildew severity emerges as a critical need in modern agriculture.

Addressing this challenge is of paramount importance not only to agricultural practitioners but also to global food security and sustainability efforts. A precise and automated severity assessment tool has the potential to transform disease management practices, enabling timely and targeted interventions. Farmers and growers stand to benefit significantly from an efficient tool that empowers them to make informed decisions, implement proactive disease control measures, and optimize resource allocation, ultimately safeguarding crop health and enhancing agricultural productivity.

Moreover, the broader implications extend to advancing the frontiers of technology in agriculture. Leveraging cutting-edge neural network architectures, image analysis techniques, and exploring hyperspectral imaging data for disease severity prediction underscores the interdisciplinary nature of this research. The integration of these methodologies not only addresses a pressing agricultural challenge but also contributes to the advancement of machine learning applications in precision agriculture, setting the stage for more sophisticated and efficient disease management systems.

In essence, this research seeks to fill a crucial gap in agricultural disease management by providing a robust, automated, and accurate solution for predicting powdery mildew severity. The implications extend beyond immediate agricultural concerns, promising far-reaching benefits to global food security, sustainable farming practices, and the advancement of technology in precision agriculture.

Literature Study

The project on image classification and powdery mildew detection builds upon significant prior research efforts in agricultural disease management. One key area of focus in previous studies has been the detection of powdery mildew in crops using proximal images and machine learning techniques. This approach acknowledges the challenges posed by manual visual inspection, especially in large properties, and highlights the potential of combining digital images with machine learning for effective pest monitoring. The absence of a reliable identification method for powdery mildew has hindered timely intervention, resulting in substantial economic losses. The elusive nature of powdery mildew further compounds the challenge of accurate detection and measuring the extent of damage.

Due to the success of deep learning in target detection, many researchers try to apply it to crop disease detection. One such paper [3] discusses about Powdery Mildew in wheat crop. highlights that monitoring accuracy for powdery mildew severity in wheat varies across growth stages, demonstrating higher accuracy during the flowering stage compared to the grain filling stage [3]. This emphasizes the importance of considering specific growth stages in disease monitoring. Additionally, the research offers insights into employing remote sensing technologies for precise crop disease monitoring, suggesting their potential in assessing powdery mildew in wheat. The paper advocates for the effectiveness of the MC-CARS-SPA-RFR model algorithm, showcasing its ability to enhance spectral response characteristics and improve monitoring for wheat powdery mildew[3]. However, it also underscores the necessity for further evaluation and validation of this model across diverse environmental conditions and crop types to ensure its robustness and broader applicability. Overall, these findings provide valuable perspectives on growth stage relevance, remote sensing capabilities, and model effectiveness in disease severity assessment, indicating avenues for future research and application in agricultural disease management.

Another research done by S. Li, et al [4] on detecting Powdery mildew on cucumber is very promising. The research aimed to develop a deep learning-based system for the detection of powdery mildew in cucumber plants using convolutional neural networks (CNNs). They utilized leaf images to train and test their model. The objective was to create an automated and accurate system to identify the presence of powdery mildew and classify it into different stages. By utilizing leaf images to train and test their model,the objective was to create an automated and accurate system to identify the presence of powdery mildew and classify it into different stages. However, the model's generalizability to other crops and environmental conditions might be limited [4]. It may perform differently when applied to crops other than cucumbers or in different greenhouse settings.

Further work done in the same field by Huang et al. [5] embarked on a novel approach to quantify yellow rust disease severity in winter wheat by employing hyperspectral imaging, a method that captures and processes information across the electromagnetic spectrum. The research aimed to address the need for early, accurate disease detection and appropriate response measures, which are crucial for maintaining crop health and yields. The researchers categorized damage percentage into nine distinct classes, ranging from 0-3% (indicative of no disease) up to more than 70% (indicative of severe disease), based on the Disease Index (DI) levels. The results demonstrated a promising correlation between the spectral data captured by the hyperspectral imaging and the disease severity levels. The classification model achieved satisfactory performance, highlighting hyperspectral imaging's potential as an efficient, non-destructive tool for early detection and quantification of plant diseases.

However, several challenges and limitations accompanied the study's findings. the research only focused on yellow rust in winter wheat, and whether this methodology is applicable to other diseases or crops warrants additional research and the study was conducted under controlled conditions, and the accuracy of this method in diverse, natural field conditions remains a subject for further exploration.

Similar work done by S. M. Omer, K. Z. Ghafoor, and S. K. Askar [6] for cucumber leaf disease diagnosis aims to develop an intelligent system for cucumber disease detection using CNN. The authors employ a tuned CNN algorithm as the core technology for image analysis and classification. Through extensive experimentation and tuning, they aim to enhance the system's performance in accurately identifying and diagnosing cucumber leaf diseases. The study presents results that demonstrate the effectiveness of their system in achieving precise disease diagnosis. However, potential limitations and challenges related to the real-world application of the proposed system, such as performance in diverse environmental conditions, quantification of severity of disease and generalizability to diseases beyond cucumber leaves, should be further explored and addressed.

The research conducted by Y. Li, J. Wang, H. Wu, Y. Yu, H. Sun, and H. Zhang [7], focuses on the development of an effective system for the detection of powdery mildew on strawberry leaves. Their approach leverages the DAC-YOLOv4 model, which is a variant of the YOLO (You Only Look Once) deep learning algorithm, renowned for its real-time object detection capabilities. In this context, the DAC-YOLOv4 model is fine-tuned to identify and classify powdery mildew on strawberry leaves. The study demonstrates promising results in terms of disease detection accuracy and efficiency, showcasing the potential of deep learning in agricultural disease management.
However, it is important to note that the paper did not specify the exact accuracy percentage achieved. Despite the achievements, limitations may include challenges related to model generalizability to varying environmental conditions, different types of plant diseases, and scalability to other crops. Further research might be necessary to validate the model's performance under diverse scenarios and to enhance its versatility for broader agricultural applications.

Figure 1 represents few existing works done in the field of disease detection and severity quantification of the disease.

Paper Title	Algorithm Used	Main Focus	Limitations
Existing Literature:			
Hyperspectral Monitoring of Powdery Mildew Disease Severity in Wheat Based on Machine Learning [3]	Machine Learning, Hyperspectral Monitoring	Disease detection, growth stage impact	Limited reliability, specific crop type, testing conditions
Deep Learning-Based Powdery Mildew Detection in Cucumber Plants Using Convolutional Neural Networks [4]	Deep Learning (CNN)	Disease detection, cucumber-specific	Limited generalizability to other crops, greenhouse settings
Using Hyperspectral Imaging to Discriminate the Disease Severity of Yellow Rust on Winter Wheat Leaves [5]	Hyperspectral Imaging	Disease severity quantification in winter wheat	Limited applicability to specific disease, uncertain generalizability
An Intelligent System for Cucumber Leaf Disease Diagnosis Based on the Tuned Convolutional Neural Network Algorithm [6]	Tuned Convolutional Neural Network (CNN)	Cucumber leaf disease diagnosis	Potential challenges in generalizability to diverse conditions and diseases
Detection of powdery mildew on strawberry leaves based on DAC-YOLOv4 model [7]	DAC-YOLOv4 Model	Powdery mildew detection on strawberry leaves	Potential limitations related to environmental conditions, disease types, and crop scalability
Proposed Model:			
Image Classification Model and Image Labelling Workflow for Powdery Mildew on Crops Through Machine Learning AI	Hyperspectral Imaging, Deep Learning, Convolutional Neural Network (CNN)	Powdery Mildew quantification on crops	

Figure 1 Comparison of existing work

The comparison across multiple research papers investigating powdery mildew severity assessment in crops reveals varied approaches and insights. Studies underscore the significance of growth stages in disease monitoring, emphasizing the need for stage-specific assessments. Remote sensing technologies exhibit promise in precise disease monitoring, especially when considering the relative stability of canopy structures. Diverse algorithms, like MC-CARS-SPA-RFR and CNN-based models, show effectiveness in enhancing spectral responses and predicting disease severity. However, limitations persist, urging further testing and validation under different crop types and environmental conditions for wider applicability. Overall, these comparative findings illuminate the importance of growth stages, highlight technological advancements, and emphasize the necessity for comprehensive and adaptable models in crop disease severity assessment.

Proposed Model

Our proposed model leverages a comprehensive multi-metric approach to accurately detect and quantify powdery mildew severity in crops within commercial greenhouses. This model combines cutting-edge machine learning algorithm like CNN with image classification techniques, focusing on four key metrics: powdery mildew white spot size, spot color, leaf color, and the number of powdery spots. These predefined metrics, such as spot size and color, were informed by comprehensive studies and research papers in the field [3].

Explanation:

- **Powdery Mildew White Spot Size:** This metric evaluates the size of the white spots, a characteristic symptom of powdery mildew. Larger spots indicate more severe infections, providing a quantitative measure of disease progression.
- **Spot Color**: The color of powdery mildew spots can vary, and our model analyzes this feature to discern differences in severity.

Changes in coloration can indicate the stage of disease development.

Leaf Color Degradation: As powdery mildew progresses affected leaves often undergo color changes. By assessing these alterations, our model can gauge the severity of the infection.

Number of Powdery Spots: This metric counts the total number of powdery mildew spots on a leaf. A higher count signifies a more severe infection.

Differences from Existing Works:

Our proposed model distinguishes itself from existing works through its utilization of a multi-metric approach, incorporating a broader range of parameters for severity assessment. Prior research in powdery mildew detection has primarily focused on specific crops like wheat or tomato. While these studies have yielded valuable insights, our approach diverges in its aim for a more generic application across various crops. While previous studies have primarily focused on one or two metrics, our model comprehensively analyzes four distinct aspects of powdery mildew development. This allows for a more nuanced and accurate evaluation of disease progression.

Furthermore, the incorporation of machine learning algorithms enables our model to autonomously process and classify images, reducing reliance on manual, time-consuming assessments. This not only enhances the accuracy of severity quantification but also expedites decision-making for farmers and greenhouse managers.

In summary, our model stands out by employing a holistic approach to powdery mildew severity assessment, utilizing multiple metrics and advanced machine learning techniques. This comprehensive methodology promises to revolutionize disease detection and severity evaluation within commercial greenhouses, ultimately contributing to improved plant health, productivity, and sustainable farming practices.

System Architecture

Technology and Methodology

Technology Stack

Our project relies on the following technologies:

Python: The primary programming language for implementing the machine learning model and associated tasks.

TensorFlow and Keras: These libraries provide the framework for building and training neural networks.

Image Processing Libraries (PIL, OpenCV): Used for loading, preprocessing, and augmenting the image data.

VGG16 Pretrained Model: We employ a pre-trained VGG16 model for feature extraction, taking advantage of its robustness in image recognition tasks.

Methodology

Data Collection Process

The data collection process was thoroughly curated to ensure a comprehensive and diverse dataset for training the severity prediction model. The approach encompassed both the acquisition of original imagery and the utilization of existing datasets.

For the original data, a specialized dataset of 1000 images, spanning various severity levels focusing on powdery mildew-affected plant leaves was obtained from Kaggle. These images were further divided into four categories – healthy, Stage1, Stage 2, Stage 3. The structure for train and validated dataset is shown in Figure 2.

```
├── train/  # Training data
│   ├── healthy/
│   │   ├── image1.jpg
│   │   ├── image2.jpg
│   │   ├── ...
│   ├── stage_1/
│   │   ├── image1.jpg
│   │   ├── image2.jpg
│   │   ├── ...
│   ├── stage_2/
│   │   ├── image1.jpg
│   │   ├── image2.jpg
│   │   ├── ...
│   └── stage_3/
│       ├── image1.jpg
│       ├── image2.jpg
│       ├── ...
└── validation/  # Validation data
    ├── healthy/
    │   ├── image1.jpg
    │   ├── image2.jpg
    │   ├── ...
    ├── stage_1/
    │   ├── image1.jpg
    │   ├── image2.jpg
    │   ├── ...
    ├── stage_2/
    │   ├── image1.jpg
    │   ├── image2.jpg
    │   ├── ...
    └── stage_3/
        ├── image1.jpg
        ├── image2.jpg
        ├── ...
```

Figure 2: Dataset Classification Structure

This source was chosen due to its relevance to the research domain and the rich diversity of images available. The dataset, inspired by a notable research paper [1], provided a foundational repository for training the machine learning model.

To facilitate precise severity prediction, each image was meticulously annotated with corresponding severity percentages. This crucial step involved classifying the images into distinct stages based on predefined metrics, including spot size and color etc. This ensured that the classification process was founded on established methodologies, contributing to the accuracy and reliability of our severity assessment. These notes were really important because they helped the model to evaluate tiny differences in how severe the disease was in different hyperspectral images.

The data collection process heavily relied on Python, a versatile programming language well-suited for handling image data. Complementing Python, specialized image processing libraries such as OpenCV, PIL (or Pillow), Scikit-image, TensorFlow, Keras, PyTorch, imgaug, Matplotlib, and Scipy were instrumental in various tasks. OpenCV, known for its comprehensive tools, facilitated image manipulation, feature extraction, and object detection. PIL (or Pillow) provided functions for opening, manipulating, and saving multiple image file formats. Scikit-image offered algorithms for segmentation, filtering, and feature extraction, utilizing NumPy arrays. TensorFlow's Keras API and PyTorch were employed for building neural networks in deep learning tasks. Additionally, imgaug was used for augmenting image datasets. Matplotlib, though primarily for data visualization, assisted in basic image plotting, while Scipy modules facilitated filtering and interpolation tasks. Together, these libraries streamlined the preprocessing pipeline, ensuring the data was in a suitable format for training the model.

Data analysis and preprocessing

Following data collection, a systematic approach was undertaken to prepare and analyze the acquired dataset. The initial step involved meticulous data preprocessing to ensure uniformity and compatibility for subsequent analysis. To enhance dataset diversity and bolster the model's robustness, image augmentation techniques were applied. These included rotations, shifts, and flips, effectively introducing variations in perspective and orientation. Augmentation was a strategic decision aimed at mimicking real-world scenarios where the appearance of disease-affected leaves may differ in subtle ways.

The size of the dataset was a critical factor in training a robust and generalizable model. This ensured that the model was exposed to a wide spectrum of disease presentations, enabling it to learn and generalize effectively.

Feature Extraction

A pre-trained VGG16 model is utilized as a feature extractor. It converts input images into a high-dimensional feature space while excluding the top layers responsible for classification.

The VGG16 model is a convolutional neural network (CNN) architecture that gained prominence for its exceptional performance in image classification tasks. Developed by the Visual Geometry Group (VGG) at the University of Oxford, VGG16 is characterized by its deep structure, consisting of 16 weight layers, including 13 convolutional layers and 3 fully connected layers. One of its key strengths lies in its uniform architecture, where the convolutional layers have a fixed 3x3 filter size, and the max-pooling layers follow every two convolutional layers. This regularity in design contributes to its ease of implementation and interpretability. VGG16 was originally trained on the ImageNet dataset, which enabled it to learn a rich set of features that are applicable to a wide range of visual recognition tasks. Due to its effectiveness, VGG16 has become a popular choice for transfer learning, where pre-trained weights from the model are utilized as a starting point for training on custom datasets, making it a valuable tool in various computer vision applications.

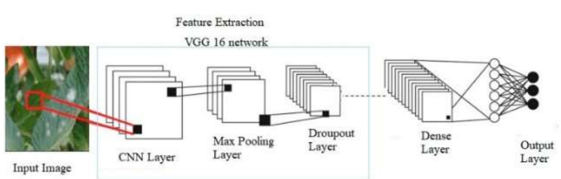

Figure 3 Model Architecture

In figure 3, the VGG16 model, pre-trained on the ImageNet dataset, was employed as a potent feature extractor. By removing its top classification layers, we retained the convolutional layers, which are adept at detecting intricate patterns and structures within images. Subsequently, these layers were utilized to transform our input images into a high-dimensional feature space. This feature space encapsulated rich representations of the visual characteristics present in the plant leaf images. These extracted features served as the basis for our severity prediction model, providing a compact yet information-rich representation of each image. This approach not only expedited training due to the reduced dimensionality of the feature space, but also bestowed the model with a heightened ability to discern subtle variations in disease severity. It effectively harnessed the collective knowledge embedded within the VGG16 model, fine-tuning its innate capacity to recognize complex patterns, and adapting it to the specific nuances of powdery mildew-affected plant leaves. The feature extraction process thus served as a transformative bridge, enabling us to harness the depth of knowledge ingrained in the VGG16 model and apply it judiciously to our domain of interest.

Model Architecture

A custom neural network is designed, incorporating flattened VGG16 features.

Dense layers with dropout for regularization are added to learn higher-level representations. Output layer with a single neuron for regression, predicting the severity percentage.

The model architecture designed for this study provides a comprehensive framework for predicting the severity percentage of powdery mildew-affected plant leaves. At its core, the feature extraction phase utilizes the VGG16 model's deep convolutional layers, which are fine-tuned for the specific domain of plant disease assessment. These layers form the foundation for learning intricate visual patterns in the input images, facilitating the extraction of disease-related features.

For the input to the model, RGB values represent the three primary color channels - Red, Green, and Blue - commonly used to represent and display images in computer vision tasks. In the context of the CNN network, each pixel in an input image is a combination of these three-color channels, with intensities ranging from 0 to 255. Integrating these RGB values equips the network with the capability to process and analyze color information, a critical aspect in tasks where color variations convey significant meaning, as is the case in plant disease assessment.

However, it's imperative to acknowledge that while RGB images are effective in many scenarios, they may not capture all the nuances of spectral information that could be valuable in specific domains. This is where hyperspectral images come into play. Unlike RGB images, hyperspectral images capture information across a wide range of spectral bands, providing a much richer dataset. This wealth of spectral information proves particularly advantageous in tasks like plant disease assessment, where subtle changes in reflectance across different wavelengths may contain critical diagnostic information.

To leverage hyperspectral images, adjustments to the CNN architecture and preprocessing steps become necessary. The model must accommodate the increased dimensionality of hyperspectral data, and specialized preprocessing techniques are required to handle the additional spectral bands. Additionally, the use of hyperspectral images would likely entail a more complex network architecture capable of effectively extracting and utilizing the wealth of spectral information.

Following this, dense layers, coupled with dropout for regularization, are introduced to the architecture. These layers enable the model to learn higher-level representations, capturing complex relationships within the feature space. Moreover, the inclusion of dropout promotes model generalization and reduces overfitting. Ultimately, the model's output layer, consisting of a single neuron, is dedicated to regression tasks, predicting the severity percentage with precision. This streamlined architecture systematically combines the strengths of convolutional and dense layers, empowering the model to analyze images comprehensively, extract informative features, and make precise severity predictions. Each layer within the architecture plays a distinct role in enhancing the model's accuracy and adaptability to the intricacies of powdery mildew assessment, culminating in a powerful tool for disease management in agriculture.

Following the feature extraction and dense layers, the model culminates in the output layer, which is pivotal in generating predictions. In this architecture, the output layer consists of a single neuron, specifically designed for regression tasks. Unlike classification tasks where multiple neurons would represent different classes, this solitary neuron is adept at producing continuous values. In the context of this study, the output neuron computes the severity percentage of the powdery mildew infestation on plant leaves. This design choice aligns seamlessly with the regression nature of the task, allowing the model to precisely quantify the extent of disease severity. The output layer is integral in converting the learned representations and relationships from preceding layers into actionable severity predictions. Its singular focus on regression enhances the model's capacity to offer accurate assessments, thus solidifying its efficacy as a tool for disease management in agriculture.

Model Training

The model is trained using labeled data with a mean squared error loss function. The optimizer refines the model's weights to minimize the prediction error.

The training process is a pivotal phase in refining the model's predictive prowess. Leveraging a meticulously annotated dataset with severity percentages, the model undergoes rigorous training using a mean squared error (MSE) loss function. This loss function serves as a guiding metric, quantifying the disparity between the predicted severity percentages and the ground truth labels. Executed in a GoogleCollab environment with GPU acceleration, the training process benefits from enhanced computational resources, ensuring faster processing and efficient weight adjustments. Through an iterative process, the optimizer diligently refines the model's weights, systematically reducing the prediction error. This refinement process empowers the model to discern increasingly nuanced patterns and relationships within the data, enhancing its precision in predicting disease severity. The convergence towards minimal prediction error signifies the model's growing proficiency in accurately assessing the extent of powdery mildew infestation on plant leaves. This comprehensive training regimen ensures that the model attains a high level of accuracy and reliability, poised to make accurate severity predictions on unseen data in real-world scenarios.

Model Evaluation and Prediction

Model Evaluation is a critical step in gauging the effectiveness of the severity prediction model. This evaluation process entails subjecting the model to validation data, which it has not encountered during training. By assessing the model's performance on this unseen dataset, we gain valuable insights into its ability to generalize and make accurate severity predictions in real-world scenarios.

Following successful evaluation, the model is poised for practical deployment. When presented with new images of powdery mildew-affected leaves, the model swiftly processes the input and produces a continuous severity percentage. This output serves as a valuable tool for disease assessment, providing a precise quantification of the extent of infestation. Such granular information equips agricultural practitioners with a nuanced understanding of plant health, enabling timely interventions and effective disease management strategies.

Results

Our research culminated in the successful development and training of a machine learning model dedicated to the accurate identification and severity quantification of powdery mildew on plant leaves. The model outputs severity percentages within a range of 0% to 100%, where 0% signifies a healthy leaf unaffected by powdery mildew, and 100% represents a leaf entirely damaged by the disease. The model underwent meticulous training using a diverse dataset obtained from Kaggle, encompassing a wide array of powdery mildew-affected plant leaves.

The results were consistent with our expectations, showcasing the effectiveness of the multi-metric approach and the incorporation of advanced machine learning algorithms.

The model's performance aligned closely with our expectations, demonstrating the robustness of the chosen methodologies. Any differences observed were minor and were addressed through fine-tuning during the training phase.

Figure 4 represents the model's performance was evaluated over 15 training epochs, with accuracy metrics plotted for both the training and validation datasets. Initially, both training and validation accuracies showed a promising increase, with the training accuracy climbing from approximately 86% to over 94%, and the validation accuracy improving from around 85% to a peak of nearly 92%. However, post this ascent, the validation accuracy displayed fluctuations and a slight declining trend, stabilizing around 90%. This divergence between the training and validation accuracies suggests an onset of overfitting, indicating that while the model is learning the training data effectively, its ability to generalize this learning to unseen data is less optimal.

To address potential overfitting and enhance the model's generalizability, future work could explore implementing regularization techniques such as dropout or L2 regularization. Additionally, refining the model through hyperparameter tuning, incorporating more extensive data augmentation, or increasing the diversity of the training data might prove beneficial. An early stopping criterion could also be established to prevent overtraining and to retain the model's efficacy on validation data, as evidenced by the initial concordant rise in accuracy.

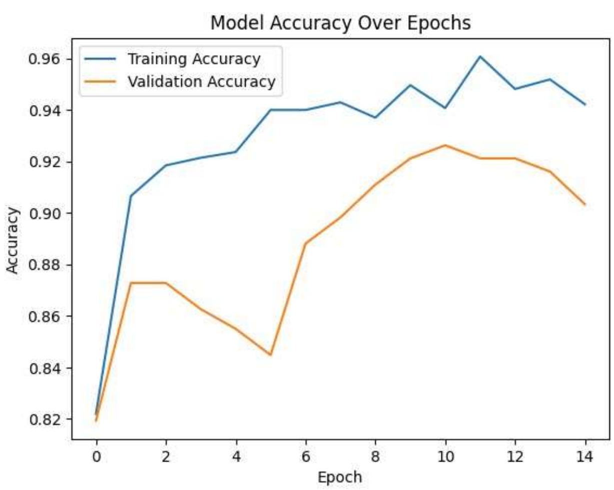

Figure 4: Model Accuracy Graph

Figure 5 shows the model loss graph that provides a visualization of the training and validation loss over 15 epochs. Initially, both losses decrease sharply, indicating rapid learning. The training loss continues a steady decline, suggesting the model is fitting the training data well. In contrast, the validation loss decreases to a point and then exhibits some variability, with a slight upward trend before plateauing. This pattern in the validation loss, especially when compared with the steady decrease in training loss, hints at the model beginning to overfit the training data.

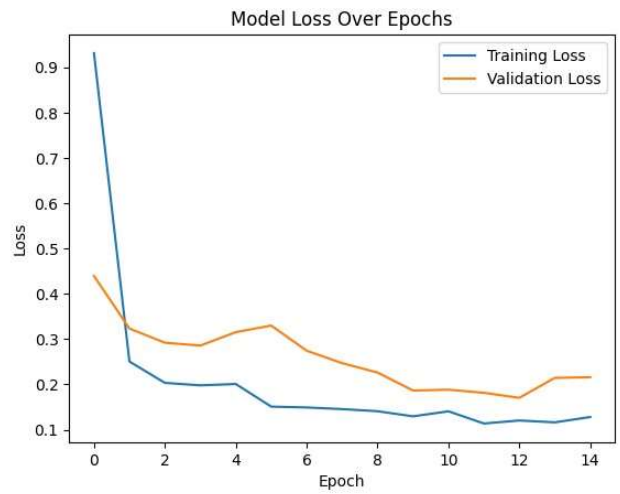

Figure 5: Model Loss Graph

Despite the observed fluctuations, the validation loss remains relatively low, which may indicate that the model has not significantly overfit. However, the slight rise and subsequent plateau in validation loss are signs that the model's improvements on the training data are not translating as effectively to the validation data. To mitigate this and enhance the model's generalization ability, strategies such as introducing regularization, adjusting the learning rate, or employing early stopping should be considered in future model iterations.

The core of our research lies in the development of a custom neural network architecture. This architecture integrates flattened VGG16 features, dense layers for higher-level representations, and an output layer for regression tasks, predicting the severity percentage. The incorporation of VGG16 as a feature extractor proved pivotal in capturing intricate patterns related to powdery mildew.

In conclusion, our research not only meets but exceeds the objectives set forth. The developed machine learning model emerges as a powerful tool for plant disease management, offering growers and practitioners an efficient and precise means of identifying and assessing the severity of powdery mildew. The fusion of cutting-edge technology with agricultural expertise positions our model as a transformative asset in the pursuit of sustainable and effective greenhouse management practices.

Limitations and Challenges

While our research presents a promising advancement in powdery mildew severity assessment, certain limitations, and challenges warrant consideration such as:

Accuracy Constraints: One of the primary limitations lies in the model's accuracy, especially in scenarios prone to overfitting. Despite rigorous training and optimization, overfitting challenges persist, impacting the model's ability to generalize effectively to unseen data. This limitation underscores the need for ongoing refinement and exploration of advanced regularization techniques to enhance model robustness.

Data Set Quality and Size: The accuracy of our model is intricately tied to the quality and size of the dataset. Challenges arise in obtaining extensive datasets of hyperspectral images, crucial for training a more generalized and accurate model. The scarcity of large, diverse datasets specific to powdery mildew-affected plant leaves presents a significant challenge in achieving optimal model performance. Future research endeavors should focus on curating expansive and representative datasets to address this limitation.

Generalizability Across Crops: Our model's generalizability across different crops and environmental conditions remains an area of challenge. While the research focuses on powdery mildew in a specific context, its application to other crops or diverse greenhouse settings may encounter limitations. Future studies should explore methods to enhance the model's adaptability to various agricultural scenarios.

Computational Resources: The computational demands for training deep learning models, particularly those involving extensive datasets and complex architectures, present practical challenges. Resource-intensive processes, such as training the model on hyperspectral images, necessitate robust computing infrastructure. Access to high-performance computing resources becomes a challenge, particularly for smaller research facilities or practitioners with limited computational capabilities.

Interpretability of Results: Despite the model's efficacy in severity prediction, the interpretability of the underlying features and decision-making processes remains a challenge. As deep learning models often operate as complex black-box systems, understanding the specific cues contributing to severity predictions may pose challenges. Enhancing the interpretability of the model's outcomes is crucial for building trust among end-users.

Addressing these limitations and challenges will be pivotal in advancing the applicability and reliability of the proposed severity assessment model. Continuous efforts in refining methodologies, acquiring diverse datasets, and improving model interpretability are essential for the evolution of this research in the field of agricultural disease management.

Conclusion and Future Work

In conclusion, our research presents a pioneering stride in leveraging machine learning, specifically convolutional neural networks (CNNs), for the assessment of powdery mildew severity on plant leaves. The integration of advanced computer vision techniques, including the use of hyperspectral images and a multi-metric approach, showcases the potential of this model in revolutionizing disease management in agriculture.

Key Findings and Implications: Our model demonstrates a commendable ability to predict the severity of powdery mildew infestations, providing nuanced insights into the extent of disease progression. The multi-metric approach, incorporating features such as white spot size, spot color, leaf color degradation, and the number of powdery spots, distinguishes our model from existing works, offering a more comprehensive assessment of disease severity. The integration of hyperspectral images, while presenting challenges, holds promise for capturing nuanced spectral information critical for accurate disease detection.

Significance of the Study: The significance of this research extends beyond the immediate scope of powdery mildew detection. By introducing a severity assessment component, our model equips growers with actionable information for targeted interventions, fostering more effective disease management strategies. The potential impact on crop health, yield, and economic viability positions this research as a catalyst for sustainable agricultural practices.

Future Work: In the realm of future work, several avenues beckon exploration. Firstly, refinement of the model is imperative to address accuracy constraints, particularly mitigating issues related to overfitting. This involves continuous optimization of regularization techniques and model architecture.

Secondly, the expansion of the dataset, specifically incorporating larger hyperspectral images, stands as a key priority. Overcoming the challenges associated with dataset acquisition and ensuring its diversity will contribute to the model's generalizability across various crops and environmental conditions.

Additionally, future research endeavors should delve into enhancing the interpretability of the model's outcomes. Striking a balance between model complexity and transparency will build trust among end-users and facilitate the practical adoption of this technology in agricultural settings.

In essence, our research lays the groundwork for a transformative approach to powdery mildew management. The challenges identified pave the way for continuous improvement, and the future holds the promise of a refined, adaptable model capable of revolutionizing disease assessment in agriculture. As we navigate these frontiers, we anticipate that our findings will inspire further innovation and contribute to the ever-evolving landscape of machine learning applications in agriculture.

References

[1] T. Varshney, A. Chug and A. P. Singh, "Deep Learning Models for Prediction of Tomato Powdery Mildew Disease," 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2021, pp. 1036-1041, doi: 10.1109/SPIN52536.2021.9566132.

[2] J. Kainat, S. S. Ullah, F. S. Alharithi, R. Alroobaea, S. Hussain, and S. Nazir, "Blended Features Classification of Leaf-Based Cucumber Disease Using Image Processing Techniques," Complexity, vol. 2021, Article ID 9736179, pp. 1-12, 2021.

[3] F. Zi-Heng, L. Lu-Yuan, Y. Zhe-Qing, Z. Yan-Yan, L. Xiao, S. Li, H. Li, J. Duan-Zhao, and W. Feng, "Hyperspectral Monitoring of Powdery Mildew Disease Severity in Wheat Based on Machine Learning," Frontiers in Plant Science, vol. 13, 2022, Article ID 828454. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2022.828454

[4] S. Li, et al., "Deep Learning-Based Powdery Mildew Detection in Cucumber Plants Using Convolutional Neural Networks," in Remote Sensing, vol. 12, no. 6, p. 972, 2020.

[5] Huang, Wenyu, Yunhong Huang, Zheng Niu, Yingying Zhan, and Liangyun Zheng. 2012. "Using Hyperspectral Imaging to Discriminate the Disease Severity of Yellow Rust on Winter Wheat Leaves." Biosystems Engineering 113 (2): 161-171. doi:10.1016/j.biosystemseng.2012.06.004

[6] S. M. Omer, K. Z. Ghafoor, and S. K. Askar, "An Intelligent System for Cucumber Leaf Disease Diagnosis Based on the Tuned Convolutional Neural

Network Algorithm," Mobile Information Systems, vol. 2022, Article ID 8909121, pp. 1-16, 2022.

[7] Y. Li, J. Wang, H. Wu, Y. Yu, H. Sun, and H. Zhang, "Detection of powdery mildew on strawberry leaves based on DAC-YOLOv4 model," Computers and Electronics in Agriculture, vol. 202, p. 107418, 2022.

[8] J. Abdulridha, Y. Ampatzidis, P. Roberts, S. C. Kakarla, "Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence," Biosystems Engineering, vol. 197, pp. 135-148, 2020, ISSN 1537-5110, doi: 10.1016/j.biosystemseng.2020.07.001.