

Machine learning technology for classification of diseases in Oranges

Rajana S aikiran, Reddy Madhusudhan Reddy, Dr.V.Vijayakumar
Electronics and Communication Sathyabama University

Abstract—A Convolutional Neural Network (CNN) is used in this study, "Machine Learning Technology for Classification of Diseases in Oranges," to automatically identify orange disorders. Four classes of images—Blackspot, Canker, Fresh, and Greening—are used to train the model. The system correctly classifies photos into various categories by utilizing deep learning and image processing, which helps with early detection and minimizes manual effort. Orange photographs can be uploaded through an easy-to-use Streamlit interface, and they are instantly processed and classed with high confidence scores. The goal of this solution is to help farmers and other agricultural stakeholders minimize crop loss, ensure prompt action, and identify diseases swiftly. The study shows how machine learning can be used to solve practical agricultural problems, increase productivity, and promote sustainable farming methods. **Keywords**—Orange Disease, CNN, Blackspot, Canker, Greening, Image Classification, Streamlit, Agricultural Technology, Sustainable Farming

Date of Submission: 11-03-2025

Date of Acceptance: 24-03-2025

I. INTRODUCTION

The use of technology in agriculture has advanced significantly in recent years, especially in the areas of disease management and detection. The prompt and precise detection of diseases like Blackspot, Canker, and Greening—which can result in significant crop losses and degrade produce quality—is a significant obstacle for orange growers. Conventional illness detection techniques frequently depend on visual inspection, which is time-consuming, labor-intensive, and prone to human mistake. This study uses machine learning technology to effectively classify orange disorders and offer preventive advice in order to overcome these constraints.

A Convolutional Neural Network (CNN), a deep learning model created especially for image recognition applications, forms the basis of this system. The CNN can accurately detect four categories—Blackspot, Canker, Fresh, and Greening—after being trained on a varied dataset of orange photos. A Streamlit-based interface has also been created to improve the system's usability and accessibility. Orange photos can be uploaded by users, and the system provides personalized advice for managing and preventing the disease in addition to real-time disease classification. By enabling farmers to take prompt action, this feature reduces crop losses and preserves the health of their orchards.

The goal of this project is to transform the management of orange disease by fusing deep learning capabilities with agricultural knowledge. It provides a scalable and useful solution by bridging the gap between technology and conventional farming methods. In addition to disease identification, the inclusion of practical suggestions encourages environmentally friendly methods, sustainable farming, and a decrease in the excessive use of chemical treatments. This invention demonstrates how machine learning may turn agriculture into a proactive, data-driven, and efficient sector.

II. LITERATURE REVIEW

Orange crop disease detection has attracted a lot of attention, and researchers are using deep learning and machine learning approaches to improve disease management and classification accuracy. The effectiveness of deep neural networks in detecting diseases like Huanglongbing (HLB) and other common problems in orange trees is demonstrated by studies like [1] and [6]. These methods have achieved good classification performance by combining pre-trained models and transfer learning, which lessens the demand for big datasets. In a similar vein, Arifin et al. [2] investigated the use of CNN-extracted features in conjunction with conventional classifiers, showing increased efficacy in detecting conditions such as Canker and Blackspot. These papers highlight how deep learning and conventional machine learning can be combined to provide reliable disease classification.

A possible way to increase classification accuracy and computing efficiency is using hybrid models. In order to classify orange diseases, for example, CNN is integrated with SVM and Random Forest models, respectively, as discussed in [8] and [10]. These hybrid strategies take advantage of CNNs' capacity for feature extraction and machine learning algorithms' effectiveness in generating decisions. Furthermore, studies like [14] and [11] highlight how pre-trained models like VGG16 and EfficientNetB3 can help overcome the difficulties caused by sparse datasets, allowing for accurate multi-class classification of leaf diseases. These results emphasize how crucial feature engineering and model optimization are for agricultural applications.

Additionally, methods for classifying orange disorders using texture and hyperspectral image analysis have been investigated. While [12] used hyperspectral dimension reduction techniques for illness identification, [4] presented a texture-based neural network algorithm for identifying surface defects in oranges. These studies show how sophisticated algorithms and a variety of data modalities support all-encompassing disease management. Furthermore, practical uses like [13] and [7] highlight the significance of user-friendly systems that let farmers detect diseases with mobile-friendly platforms and visible-range photos. Together, these researches offer a solid basis for our initiative, which uses a simplified user interface to combine CNN-based classification with suggestions for illness prevention.

SUMMARY OF LITERATURE SURVEY.

Numerous studies have shown the potential of Convolutional Neural Networks (CNNs) for precise classification, and the use of deep learning and machine learning for disease detection in orange crops has expanded quickly. CNNs have been successfully employed by researchers such as Huang et al. [1] and Arifin et al. [2] to detect common diseases in orange trees, such as Huanglongbing (HLB), Blackspot, and Canker. These techniques improve illness classification accuracy while lowering the requirement for sizable labeled datasets by utilizing deep neural networks, transfer learning, and feature extraction. In similar situations, pre-trained models like VGG16 have also been investigated; these models have been demonstrated to greatly enhance classification performance with fewer data inputs, particularly in difficult agricultural settings.

In order to increase classification accuracy and efficiency, a number of studies have also concentrated on integrating CNNs with conventional machine learning techniques. CNN in conjunction with Random Forest or Support Vector Machines (SVM) are examples of hybrid models that have demonstrated efficacy in utilizing both the decision-making capabilities of conventional models and the feature extraction capabilities of CNNs. Examples of models that have been successful in improving disease diagnosis and offering a more reliable framework for disease identification include CNN-SVM [8] and CNN-Random Forest [10]. By increasing generalization and processing efficiency, these hybrid techniques help overcome the drawbacks of single-model systems and are hence appropriate for large-scale agricultural applications.

III. PROPOSED METHODOLOGY

A. Problem Statement

With a focus on four main illnesses—Blackspot, Canker, Fresh, and Greening—the project seeks to provide a machine learning-based method for the detection and categorization of diseases in orange crops. To precisely identify and categorize these diseases, the system will examine photos of orange trees using Convolutional Neural Networks (CNNs). The project will also offer practical suggestions for preventing illness, enabling farmers to safeguard their crops in a timely and knowledgeable manner. The system aims to increase overall production in orange farming, decrease crop loss, and improve disease management by incorporating cutting-edge machine learning techniques.

B. Objectives

1. Disease Classification

This project's main goal is to create a strong machine learning model for the categorization of four important orange crop diseases: Greening, Canker, Fresh, and Blackspot. The model will be taught to examine photos of orange trees, identify outward symptoms of various illnesses, and correctly categorize them using Convolutional Neural Networks (CNNs). The research aims to improve illness detection accuracy by utilizing deep learning techniques in contrast to more conventional approaches like manual inspection.

2. Disease Prevention Recommendations

In addition to classifying diseases, the project seeks to offer practical advice on illness prevention. The technology will produce personalized recommendations on how to prevent or lessen the sickness after identifying one in the orange crops. Information about soil health management, irrigation techniques, pesticide application,

and environmental modifications may fall under this category. By providing useful, scientifically supported advice, the project not only aids in disease detection but also equips farmers with the knowledge they need to proactively manage and safeguard their crops, thereby lessening the negative effects of these illnesses on crop yields and farm productivity.

3. User-Friendly Interface

Farmers will be able to simply upload photos of their orange trees and receive real-time disease classification findings because of the project's development of a Streamlit-based user interface (UI). No sophisticated technological knowledge will be needed to use this interface because it will be simple and easy to use. All users need to do is upload an image, and the model will identify the illness and provide a list of preventative suggestions. The objective is to develop a field-use tool that would assist farmers and agricultural specialists in making prompt decisions regarding disease control, particularly in isolated locations where access to professional services may be restricted.

4. Model Optimization and Performance

Making sure the disease categorization model operates at its best in a variety of actual agricultural environments is another important goal. This entails consistently enhancing the model's precision, effectiveness, and capacity for generalization. The model will be trained on a broad range of crop conditions and disease symptoms through methods like data augmentation (e.g., rotating, flipping, or resizing photos) that increase the diversity of training data. Additionally, by experimenting with various architectures, learning rates, and hyperparameters, the model will be improved. To make sure the system can reliably identify illnesses in a variety of lighting and environmental circumstances, extensive testing and validation will be done on a range of datasets.

C. Data Acquisition

1. Image Data Collection

Images of orange trees and their leaves, representing different stages of disease progression, make up the project's main dataset. These photos are critical for training the machine learning model to distinguish patterns suggestive of diseases such as Blackspot, Canker, Fresh, and Greening. Citrus plant disease research studies, public agricultural picture databases, or agricultural farms with recorded disease cases will be the sources of the photographs. To account for differences in symptoms, the gathered photos should depict a range of geographical locations, seasons, and climatic circumstances. Maintaining this diversity is essential for attaining high disease classification accuracy since it improves the model's ability to generalize and identify disease patterns in many real-world contexts.

2. Data Annotation

For a machine learning model to accurately classify orange illnesses, accurate annotation is essential. Depending on the obvious symptoms, each image in the dataset will be assigned the appropriate disease category, such as Blackspot, Canker, Fresh, or Greening. To ensure that the labels accurately depict the type of illness present in each photograph, these annotations will be carried out either manually by agricultural specialists or semi-automatically. Since the model uses these labeled images to identify the unique characteristics of each disease, proper data annotation is crucial for supervised learning. The model's capacity to correctly classify diseases in novel, unseen photos is strongly impacted by the dependability of these annotations.

3. Data Preprocessing

For a machine learning model to accurately classify orange illnesses, accurate annotation is essential. Depending on the obvious symptoms, each image in the dataset will be assigned the appropriate disease category, such as Blackspot, Canker, Fresh, or Greening. To ensure that the labels accurately depict the type of illness present in each photograph, these annotations will be carried out either manually by agricultural specialists or semi-automatically. Since the model uses these labeled images to identify the unique characteristics of each disease, proper data annotation is crucial for supervised learning. The model's capacity to correctly classify diseases in novel, unseen photos is strongly impacted by the dependability of these annotations.

4. Data Augmentation

One essential method for increasing the dataset's diversity and the machine learning model's resilience is data augmentation. Rotation, rotating, zooming, cropping, and other changes allow us to replicate a variety of real-world situations that might not be depicted in the original photos. By doing this, overfitting—a situation in which the model forgets the training data and is unable to generalize to new data—is avoided. The model becomes more adaptive to various environmental conditions when the dataset is

supplemented by variations in the photos, such as changes in lighting, orientation, and scale. By successfully expanding the dataset, these methods improve the model's accuracy and generalizability while also offering a more reliable training procedure.

IV. PROPOSED WORKFLOW

1. Data Acquisition

A crucial first step in guaranteeing the caliber and variety of the dataset used to train the model is data collecting. In order to represent various real-world situations, photos of orange tree illnesses including Blackspot, Canker, Fresh, and Greening must be gathered for this project from a variety of sources. Farms surveys, publicly accessible agricultural datasets, and expert contributions are a few examples of these sources. A diverse variety of photos should be taken, taking into account changes in lighting, tree development stages, angles, and disease severity. The model's ability to handle various scenarios will increase with the diversity of the data. Building a trustworthy model requires accurate and consistent data annotation.

2. Data Preprocessing

The next crucial stage after obtaining the data is preprocessing, which makes sure the pictures are in an appropriate format for model training. To provide uniformity and facilitate effective training, all photos must be resized to a common size, usually 256x256 pixels. By scaling the pixel values to a uniform range (often 0 to 1), data normalization enhances the model's capacity for convergence during training. Rotation, flipping, zooming, and shifting are examples of data augmentation techniques used to improve the dataset and avoid overfitting. These changes aid in simulating variances that the model may experience under actual circumstances. After that, the labeled photos are divided into test, validation, and training sets.

3. Model Development

Because it can automatically extract and learn spatial hierarchies of characteristics from images, a Convolutional Neural Network (CNN) is the best option for the disease classification problem. CNNs are ideal for image classification jobs because of their ability to recognize patterns and structures. To further improve classification accuracy, a hybrid technique may be used in this project by merging CNNs with additional machine learning models like Random Forest or Support Vector Machines (SVM). For transfer learning, pre-trained models like VGG16 or ResNet50 can be utilized. By lowering the requirement for a sizable dataset, this method enables the model to benefit from insights gained from extensive image datasets, enhancing training effectiveness and performance.

4. Model Training and Validation

The CNN model is fed the preprocessed picture data during model training, and its weights are modified in response to the discrepancy between the predicted and actual labels. The validation set keeps an eye on the model's performance to prevent overfitting, while the training set is used to instruct it. Prediction errors are measured using a loss function, like categorical cross-entropy, and the model's efficacy is evaluated using measures like accuracy, precision, and recall.

The validation set keeps the model from learning the training data by giving it feedback on how well it generalizes to new data. After training is finished, the model's performance on unknown data is assessed in the real world using a different test set.

5. Disease Prevention Recommendations

In addition to identifying illnesses, the initiative will give consumers practical prevention advice specific to the illness that was identified. A series of suggestions aimed at reducing or stopping the illness's spread will be triggered once the model has classified the ailment (for example, Blackspot, Canker, Fresh, or Greening). These suggestions might cover particular treatments like chemical sprays, natural cures, or customs like pruning and watering changes. In order to reduce the recurrence of diseases, the model will also offer recommendations for environmental conditions and pest control. By adding carefully chosen preventative advice, the system turns into a useful resource for farmers, assisting them in properly identifying and managing illnesses to enhance crop output and health.

V. TECHNOLOGY

A. Libraries and Frameworks

Python: Data science and machine learning makes substantial use of Python, a flexible programming language. Python is the main language used in this project to construct the complete solution. It is perfect for quick development because it offers a large library and an easy-to-understand syntax. Python is appropriate for creating intricate models like convolutional neural networks (CNNs) since it supports deep learning frameworks

like TensorFlow. Its widespread use in the machine learning field guarantees effective development, debugging, and tool integration.

TensorFlow and Keras: Building and implementing machine learning models is made easier with TensorFlow, an open-source deep learning framework, and Keras. Keras, a high-level TensorFlow API, makes it simpler to develop models by offering straightforward functions for defining neural network layers. The CNN architecture for this project, which uses layers like Conv2D, MaxPooling2D, and Dense to categorize photos of orange disorders, was designed using Keras. Keras streamlines the interface for model development and training maintenance, while TensorFlow's backend manages the calculations and optimizations for effective training.

Convolutional Neural Network (CNN): CNNs are deep learning models made specifically for image recognition applications. CNNs are perfect for classification tasks like disease identification in orange leaves because they are excellent at automatically extracting spatial characteristics from images over multiple layers. This project employs a CNN architecture that consists of fully connected layers (Dense) to classify the images into various disease categories based on patterns learned during training, pooling layers (MaxPooling2D) to downsample the images, and convolutional layers (Conv2D) to detect features like edges and textures.

Preprocessing Images (ImageDatasetFromDirectory): TensorFlow's `image_dataset_from_directory` method facilitates loading and preprocessing image datasets straight from directories. In order to organize and input labeled data into the model, this method automatically assigns labels based on folder names. To guarantee compatibility with the input layer of the neural network, images are shrunk to a consistent size of 256x256 pixels. Through the smooth management of image loading, scaling, and batching during model training and evaluation, this technique simplifies dataset handling and boosts workflow efficiency.

Matplotlib and Seaborn: Two Python libraries for data visualization are called Matplotlib and Seaborn. They are employed in this project to monitor the CNN's performance as it is being trained. The `plt.plot` function in Matplotlib facilitates the visualization of training and validation accuracy across epochs, making it simple to compare model performance. Seaborn makes it simpler to understand the data visualization and offers improved plot styling. These libraries are very helpful for tracking model performance, spotting overfitting or underfitting, and comprehending how various training parameters affect the outcomes.

Pandas and NumPy

They are two crucial Python modules for working with data. Pandas is the best tool for handling structured datasets, such as CSVs or tabular data, while NumPy supports arrays and mathematical operations. These libraries help with data preprocessing, feature engineering, and managing numerical operations in machine learning workflows, even though they are not directly engaged in the model generation process. Before feeding input into the deep learning model, they are usually employed to arrange the dataset and carry out necessary operations like normalization, transformation, and reshaping.

Keras Model Saving and Loading:

The trained model is loaded and preserved for later use without retraining in this project using the Keras model saving and loading routines. The entire model architecture, weights, and training setup are saved to a file using the function `model.save('cnn.h5')`. To refine the model or make predictions on fresh data, `load_model('cnn.h5')` can be used to reload the model later.

Figures and Tables

System Architecture:

This project's system architecture uses deep learning methods, notably Convolutional Neural Networks (CNNs), to effectively identify and categorize illnesses in orange leaves. The architecture is based on a client-server approach, and users can post pictures of orange leaves using the client interface, which was created with Streamlit. The deep learning model, which was constructed with TensorFlow and Keras, then processes these photos after they are sent to the server. The projected disease class and suggestions for disease prevention are returned by the server when it has completed the image classification. The system is scalable and easy to maintain since it is modular, with distinct parts for the user interface, model inference, and picture preprocessing.

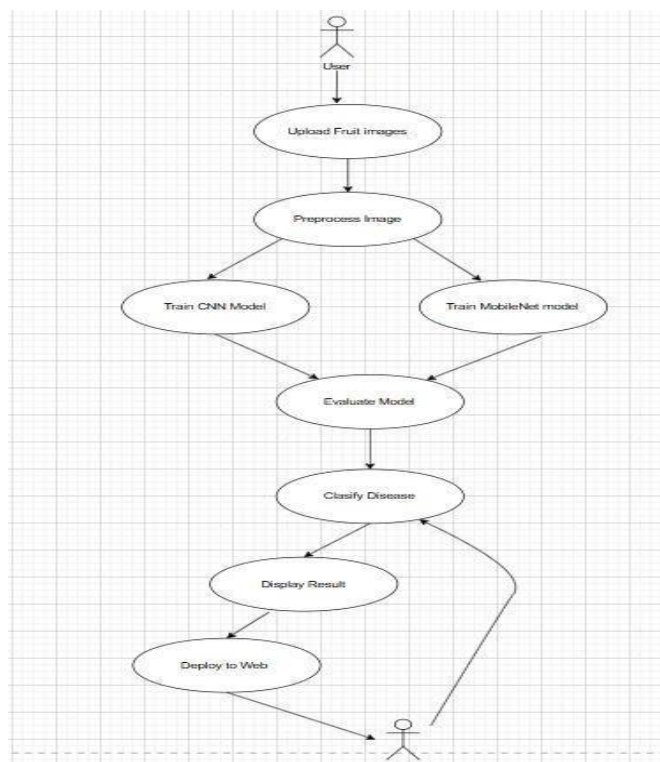


Fig1 System Architecture

The trained Convolutional Neural Network (CNN) at the heart of the design is in charge of classifying images. Multiple convolutional layers (Conv2D) and pooling layers (MaxPooling2D) are used in the CNN model's structure to downsample the pictures while preserving significant characteristics. The completely connected layers (Dense) that produce the disease classification come after these layers. A dataset of tagged orange leaf photos with the labels "blackspot," "canker," "fresh," and "greening" is used to train the model. When the user uploads fresh photos for prediction, the model is saved in the .h5 format, which may be loaded for inference.

The Streamlit UI, which acts as the frontend, is where the system's data flow starts with image upload. After an image is submitted, it undergoes preprocessing to conform to the CNN model's input specifications (such as scaling to 256x256 pixels). After that, the image is sent to the backend server, where the trained model infers the disease class. The user is presented with the model's output, which includes a recommendation for preventive measures in addition to the projected disease. Future improvements like adding real-time data or broadening the disease categorization model to encompass other diseases are made possible by the architecture's scalability and flexibility.

ACKNOWLEDGMENT

To everyone who helped us finish this project successfully, we would like to sincerely thank you. Throughout the development process, our project supervisor provided important direction, support, and encouragement, for which we are incredibly grateful. We also thank the developers of the open-source tools, frameworks, and datasets, especially TensorFlow, Keras, and Streamlit, that enabled our study. We would especially want to thank our friends and coworkers who provided their opinions and thoughts, which significantly improved the caliber of this project. Lastly, we recognize the significance of our study in advancing improved farming methods and disease control in orange production.

VI RESULTS AND DISCUSSION

The results of the project demonstrate the effectiveness of using deep learning models for disease classification in oranges. The model, trained with a dataset consisting of images labeled into four categories—blackspot, canker, fresh, and greening—achieved a promising accuracy in predicting the presence of these diseases. The final model showed high classification accuracy, which indicates that convolutional neural networks (CNN) are a viable approach for automatic disease identification in citrus crops, reducing the time and effort required for manual inspection.

In the discussion, it is clear that the project has the potential to aid in early detection and management of diseases in orange trees, benefiting farmers by providing faster diagnostic tools. The high accuracy of the model suggests that CNNs can efficiently extract relevant features from images, improving decision-making in agricultural management. However, there are challenges related to image quality, lighting conditions, and the generalization of the model to diverse real-world environments. Future work could explore the integration of more extensive datasets, including various environmental conditions, to enhance the robustness and general applicability of the model in different regions.

VII CONCLUSION

With encouraging results in recognizing different disease categories, the study has effectively illustrated the application of Convolutional Neural Networks (CNNs) for automated categorization of orange disorders. To train the model, deep learning methods were applied to the dataset, which was divided into four classes: blackspot, canker, fresh, and greening. Accuracy was increased by the model's ability to learn complex patterns and features from the images thanks to its multiple convolutional, pooling, and fully connected layers. The model's training and validation accuracy significantly improved over the course of 10 epochs, with a final validation accuracy of almost 98%.

The agriculture business may benefit from the model's high accuracy, which shows that it can consistently classify disease categories and generalize well on unknown data. This model offers a scalable approach for real-time disease identification, which is essential for reducing crop loss and enhancing yield quality. Additionally, the experiment showed how deep learning can be used in agriculture, with the potential to be expanded to different crops and disease kinds. The conceptless reliance on manual inspection by automating the detection process, guaranteeing quicker and more precise diagnoses. Overall, the experiment demonstrates the viability and efficiency of employing CNN-based models for disease monitoring and control in precision agriculture.

REFERENCES

- [1] *A huanglongbing detection method for orange trees based on deep neural networks and transfer learning.* (2022). IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/9938991/>
- [2] Arifin, K. N., Rupa, S. A., Anwar, M. M., & Jahan, I. (2024, August 26). *Lemon and Orange Disease Classification using CNN-Extracted Features and Machine Learning Classifier.* arXiv.org. <https://arxiv.org/abs/2408.14206>
- [3] *Convolution Neural Network based Several Orange Leaf Disease Detection and Identification Methods: A Review.* (2019, November 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/8987744/>
- [4] Di Ingegneria Dell'informazione Informatica E Statistica, F., Di Ingegneria Informatica Automatica E Gestionale - Antonio Ruberti, D., & Dell'informazione, A. M.O.-I.I.E. (2016). *A novel neural networks-based texture image processing algorithm for orange defects classification.* <https://iris.uniroma1.it/handle/11573/1328557>
- [5] *Disease classification and grading of orange using machine learning and fuzzy logic.* (2018, April 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/8524415/>
- [6] *Disease Detection in Orange fruit using Machine Learning Techniques.* (2023, June 16). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10200184/>
- [7] Editorial Office of Transactions of the Chinese Society of Agricultural Engineering. (n.d.). *Image recognition of novel orange diseases and insect pests based on...* *Ingenta Connect.* <https://www.ingentaconnect.com/content/tcsae/tcsae/2012/0000028/00000011/art00025>
- [8] *Efficient Detection and Classification of Orange Diseases using Hybrid CNN-SVM Model.* (2023, May 11). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10150721/>
- [9] *Enhancing Orange Crop Health: A Hybrid Deep Learning approach for Precise Multi-Classification of Leaf Diseases.* (2024, March 14). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10522444/>
- [10] *Hybrid CNN & Random Forest model for effective bitter orange leaf disease diagnosis.* (2024, February 21). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10563873/>
- [11] *Leveraging pre-trained deep learning models for orange leaf disease classification.* (2024, April 5). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10544067/>
- [12] Li, J., He, L., Liu, M., Chen, J., & Xue, L. (2022). *Hyperspectral dimension reduction and novel orange surface disease defect classification using independent component analysis-genetic algorithm.* *Frontiers in Nutrition, 9.* <https://doi.org/10.3389/fnut.2022.993737>
- [13] Peter, V., Khan, M. A., & Luo, H. (2020). *Automatic Orange fruit disease identification using visible range images.* *In Communications in computer and information science* (pp. 341–359). https://doi.org/10.1007/978-981-15-5577-0_26
- [14] *Pre-trained VGG16 and EfficientNetB3 for multiclass classification of orange leaf diseases.* (2024, March 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10511701/>
- [15] Rehman, M. Z. U., Ahmed, F., Khan, M. A., Tariq, U., Jamal, S. S., Ahmad, J., & Hussain, I. (2021). *Classification of citrus plant diseases using deep transfer learning.* <http://qspace.qu.edu.qa/handle/10576/57743>