Machine Learning-based Citrus Plant Disease Detection and Management System using Computer Vision

A.Srinivasa Revanth Krishna
Computer Science and Engineering
Hindustan Institute of Technology and Science
19113128@student.hindustanuniv.ac.in

T.Sri Siva Sandeep Reddy
Computer Science and Engineering
Hindustan Institute of Technology and Science
19113129@student.hindustanuniv.ac.in

R. Dheepthi
Ass.Professor ,Computer Science and Engineering
Hindustan Institute of Technology and Science
dheepsraj@gmail.com

***Abstract*—This article presents a novel method for identifying and managing diseases that affect citrus fruits by utilising cutting-edge computer vision and machine learning techniques. The method is presented in this article. The proposed system utilises a Convolutional Neural Network (CNN) model to extract features from images and classify them as healthy or diseased. Following this, a Random Forest algorithm is used to make the final prediction based on the extracted features. The system enables farmers to easily upload images of their citrus fruits from mobile or web-based platforms, enabling instant diagnosis of diseases and providing effective management plans to address the issue.**

The design of the system is centred on enhancing the performance of the model in recognising patterns that are associated with healthy and diseased fruits. In addition to this, it features an automatic management plan generation and alert system, both of which are designed to prompt farmers to take prompt actions for the prevention and control of diseases.
The purpose of the proposed system is to deliver an instrument that is both effective and simple to use for the diagnosis and treatment of diseases, which will ultimately result in increased crop yields and increased profitability for citrus growers. The agricultural sector as a whole stands to gain significantly from the successful implementation of this system, which has the potential to bring about significant improvements.

Keywords:Citrus fruits,Disease identification,Disease management Computer vision,Machine learning Convolutional Neural Network (CNN),Random Forest Image classification,Automatic management plan generation,Alert system.

I. Introduction

Citrus fruits are extremely valuable agricultural commodities all over the world. Some of the most common citrus fruits include oranges, lemons, and grapefruits. Citrus crops, on the other hand, are susceptible to a wide variety of diseases, which can significantly cut yields and negatively affect the quality of the fruit. The early detection and management of these diseases are essential for reducing the amount of crop loss and increasing the amount of crop production.

Manual inspections carried out by knowledgeable specialists are required by the conventional approach to disease identification and control in citrus fruits. On the other hand, these approaches require a great deal of time and effort in addition to potentially high costs. In addition, manual inspections might not be sensitive enough to detect diseases in their early stages or symptoms that are subtle.
There is a growing interest in developing automated systems for disease identification and management in citrus fruits using advanced computer vision and machine learning techniques. This is being done as a means of overcoming the challenges that have been presented. These systems are able to make a diagnosis of a disease in a crop both quickly and accurately, which enables timely management and treatment.

Using computer vision and machine learning, this paper makes a case for the development of an automated system for the detection and management of diseases affecting citrus fruits. A Convolutional Neural Network (CNN) model is used by the system to analyse images, extract features from those images, and then classify the images as healthy or diseased. After that, an algorithm called Random Forest is used by the system to make a final prediction based on the features that were extracted. The system that is being proposed enables farmers to upload images of their citrus fruits from mobile or web-based platforms, which enables instant diagnosis of diseases and provides effective management plans to address the issue.

The purpose of the system is to enhance the performance of the model by improving its ability to recognise patterns that are associated with healthy and diseased fruits. In addition to that, it features an automatic management plan generation and alert system, both of which are designed to prompt farmers to take prompt actions for the prevention and control of diseases.

The purpose of the proposed system is to deliver an instrument that is both effective and easy to use for the diagnosis and treatment of diseases, which will ultimately result in higher crop yields and greater profitability for citrus growers. The agricultural sector as a whole stands to gain significantly from the successful implementation of this system, which offers numerous advantages.

The following outline describes the format of this article. A discussion of works that are connected to this topic kicks off in the second section. In Section 3, we have explained our problem statement. In Section 4, we provide an overview of the dataset, methodolohy discussed how the data will be preprocessed, and introduce the proposed CNN Model and Random Forest design. In Section 5, we carry out all of the experiments, discuss the shortcomings of the proposed model for deep learning, and look ahead to the work that will be done in the future. Section 6 of this paper is where everything is summed up and concluded.

II. Related Work

[1] An Automatic Method for the Detection of Diseases in Citrus Fruit and Leaves Using a Deep Neural Network Model (2021) Khattak, Asad and Asghar, Muhammad Usama and Batool, Ulfat and Asghar, Muhammad Zubair and Ullah, Hayat and Al-Rakhami, Mabrook and Gumaei, Abdu: A deep learning approach that makes use of a Convolutional Neural Network (CNN) model is proposed in this paper as a means of diagnosing five common diseases that affect citrus fruit and leaves. With an accuracy of 94.55percent on the test, the model performs significantly better than other cutting-edge deep learning models. The model has a high level of accuracy, integrates multiple layers, and performs better than comparable models. These advantages set it apart from competitors. However, the model can only be applied to certain diseases, it calls for a substantial amount of data, and it is difficult to understand.

[2] Muthu Brindha, G. and Karishma, K.K. and Nivetha, J. and Vidhya, B. Automatic Detection of Citrus Fruit Diseases Using MIB Classifier (2022):Muthu Brindha, G. and Karishma, K.K. and Vidhya, B.A deep learning approach utilising a MIB classifier model is proposed in this paper as a means of diagnosing five prevalent diseases that can affect citrus fruits. The Gaussian feature extraction method is utilised in the development of this model, which results in an accuracy rating of 98percent. A high level of accuracy, efficient feature extraction, and superior performance to that of comparable models are among the benefits offered by the proposed solution. However, the model can only be applied to certain diseases, it calls for a substantial amount of data, and it is difficult to understand.

[3]Disease Detection in Fruits through the Use of Image Processing (2021): S. Malathy, R. R. Karthiga, K. Swetha, and G. Preethi. In this paper, a deep learning approach utilising CNN is proposed for the purpose of diagnosing diseases in fruits. Python is the programming language that is used to implement the model, which is so accurate that it achieves a level of 97percent. The model's strengths include the high accuracy it provides, the approach it takes, which is based on deep learning, and the possibility that it will assist farmers in improving crop growth. However, the model can only be applied to certain diseases, it calls for a substantial amount of data, and it is difficult to understand.

[4]H. Mohinani, V. Chugh, S. Kaw, O. Yerawar, and I. Dokare. Vegetable and Fruit Leaf Diseases Detection Using ResNet (2022). The ResNet algorithm is used throughout the paper to propose a solution for the detection of diseases in vegetables and fruits. The purpose of this project is to categorise plant leaves into either 14 healthy classes or 26 diseased classes. The accuracy of the test, as obtained, is 99.2 percent. The solution has a high level of accuracy, a large dataset, and the potential to improve agricultural production. These are the advantages that it offers. However, the solution can only be applied to a limited number of diseases, the ResNet algorithm can be difficult to understand, and the model is dependent on a substantial amount of data.

[5]Apple Fruit Disease Detection for Hydroponic Plants Using Cutting-Edge Technology, Machine Learning, and Image Processing (2021): K. Lisha Kamala and S. Anna Alex. The purpose of this paper is to discuss the use of image processing and machine learning techniques as a means of diagnosing diseases that may be present in apple fruits that have been grown through hydroponic farming. The objective is to detect diseases in their earliest stages in order to forestall a reduction in productivity, quality, and quantity. The approach has a number of benefits, including the early detection of diseases, improvements in both quality and quantity, and a reduction in the loss of productivity. Nevertheless, the paper does not report on the accuracy of the proposed approach, and it is possible that it is only applicable to a select number of diseases.

[6]The Identifying and Categorizing of Fruit-Related Illnesses Using Deep Learning Model, M. A. Matboli and A. Atia 2022, the objective is to use a deep learning model to identify fruit diseases in apples and citrus in order to facilitate the early detection and prompt treatment of these conditions. In this project, five different transfer learning models are evaluated, and a specialised CNN model that achieves an accuracy level of 99.16percent is chosen as the winner. The early detection of fruit diseases, which leads to timely treatment and the prevention of further spread,

is one of the advantages of this project. Other advantages include the reduction in the amount of data and resources required for training, the high accuracy of the customised CNN model, which leads to better decision-making for farmers and fruit producers, and the reduction in the amount of time spent on training. The performance of transfer learning models may be limited by the quality and diversity of the training data; the cost of implementing the proposed solution and ongoing maintenance may be significant for some fruit producers; transfer learning models may not perform well for new or rare diseases not represented in the training data; the performance of transfer learning models may be limited by the quality and diversity of the training data.

[7] An Application of Graph Convolution Neural Networks for the Detection and Classification of Banana Fruit Diseases (GCNN), The research conducted by P. Sajitha and A. Diana Andrushia, 2022.2, focuses on identifying and classifying diseases that can affect banana fruit by utilizing an innovative graph convolutional neural network. The research investigates the problem of fruit diseases that reduce crop yields and cut into farmers' profits in Asian countries where bananas are grown. The benefits of this project include improved accuracy in the detection of diseases and the categorization of banana fruits, better results compared to those obtained by other neural networks, and the overcoming of challenges associated with phytopathologyorgeneralize. The limited scope of application, the high computational power and resources required for the neural network training and testing process, and the possibility that the results will not generalise to other kinds of fruits and diseases are some of the disadvantages of this method.

[8]Fruit Disease Classification using Convolutional Neural Network,N, Pradheep and G, Praveen Raj K and Chanduru N M, Purna and N, Kalaivani and V, Nandalal,2022 seeks to classify fruits utilising a convolutional neural network with improved accuracy through the utilisation of data-enhancing methods and a category filter. Oranges were used in the experiment, and the database that was used to find images contains pictures of three different kinds of fruits. The benefits of this project include improved accuracy brought about by the application of data-enhancing methods and category filters, the specification of a single domain in order to achieve better results, and improved prediction and analysis capabilities brought about by the application of methods of more in-depth learning. The need to identify a greater variety of fruits results in an increase in the cost of the calculation, and despite efforts to eliminate potential sources of error, there is still a possibility that the accuracy will suffer as a result.

Overall, the research that has been done so far suggests that deep learning-based methods could be used to find diseases in fruits and vegetables. But these methods might only work for some diseases; they might require a lot of data; and they might be hard to put into place. The benefits of utilizing these methodologies include their high degree of accuracy as well as their potential to enhance agricultural production.

III. Problem Statement

Many farmers all over the world make a significant portion of their annual income off of the sale of citrus fruits. Citrus crops, on the other hand, are notoriously susceptible to a wide range of diseases, any one of which has the potential to cause significant damage to the harvest or even wipe it out entirely. The process of recognising and treating these diseases is one that is not only difficult but also time-consuming, and it requires specialised knowledge and years of experience. In addition, even relatively short lag times in disease diagnosis can result in significant crop losses, which in turn can cause a decrease in profits and increased financial strain for farmers. In this paper, we propose an automated system that makes use of advanced computer vision and machine learning techniques in order to diagnose and manage diseases that affect citrus fruits. This will help address the challenges that have been outlined above. The system is intended to provide farmers with an effective and user-friendly tool that will enable them to quickly identify diseases and receive management plans. This will enable the farmers to take immediate actions to prevent and control the spread of diseases. This system aims to increase crop yields and profitability for citrus growers by providing an effective solution for disease identification and management. Additionally, it aims to minimise the negative impact of diseases on the agricultural industry as a whole.

IV. Methodology

Collection of Data: The first step in the methodology is to gather together a dataset of images of citrus fruits that is representative of the whole. Images of healthy as well as diseased fruits, depicting each stage of the ailment, ought to be incorporated into this dataset. For the purpose of ensuring accuracy, the photographs should be of a high quality, taken with the correct camera settings, and under lighting conditions that are constant.

Preprocessing the Data Following the collection of the dataset, the images will need to be preprocessed in order to guarantee that the data is error-free and prepared for analysis. The performance of the machine learning model can be helped along by preprocessing techniques such as resizing, normalisation, and image enhancement. These techniques are included in preprocessing.

Convolutional Neural Network (CNN) Model Development: The next step is to develop a CNN model that can extract features from the images and classify them as either healthy or diseased. In order to obtain a high level of accuracy, the model ought to be improved by means of hyperparameter tuning.

Model Training: Once the CNN model has been developed, it needs to be trained on the preprocessed dataset in order to function properly. During the training phase, the images are loaded into the model, the parameters of the model are tweaked based on the feedback, and the process is then repeated until the model achieves an acceptable level of accuracy.

Random Forest Algorithm: Following the training of the CNN model, the system uses a Random Forest algorithm to make a final prediction based on the features that were extracted. In order to arrive at a definitive classification, the Random Forest algorithm compiles the findings and conclusions from a number of different decision trees.

Generation of Management Plans: Another component of the system under consideration is an automatic mechanism for the generation of management plans. Farmers are given a list of recommended courses of action to take in order to address the problem after the management plan, which is generated based on the diagnosis of the disease, is provided to them.

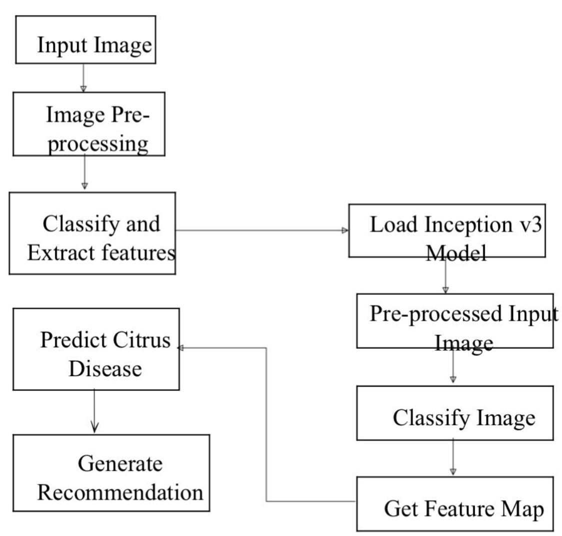

Fig. 1. Flowchart of methodology

The system not only generates management plans, but it also has an alert system that notifies farmers whenever a new disease is discovered in their crops. This system is in addition to the management plan generation that the system performs. Farmers are given the ability to take immediate action thanks to this prompt notification, which helps to prevent the further spread of the disease.

Evaluation of the System In the final step, the system that has been proposed is evaluated using a variety of evaluation metrics in order to determine its accuracy, precision, recall, and F1-score. The performance of the system is evaluated in comparison to a dataset known as the holdout, which was not utilised during the phase in which the model was being trained. The results of the evaluation are utilised to determine which aspects require additional enhancement and optimization.

A. *Hybrid Modeling*

Combining CNN and Random Forest Algorithms for Accurate Diagnosis of Citrus Fruit Diseases Using Hybrid Modeling:A hybrid modelling approach will be formed by our proposed system, which makes use of the power of both Convolutional Neural Network (CNN) and Random Forest algorithms. This approach will be used to identify and manage diseases that affect citrus fruits. Because of this approach, we are able to benefit from the qualities that are unique to each model, which ultimately leads to a diagnosis that is more precise and trustworthy.

There is no requirement for manual feature engineering because the CNN model is capable of automatically extracting features from the images. Because the model is able to recognise patterns and characteristics that human observers of the data may miss, this method results in a diagnosis of the disease that is more precise and trustworthy. In addition, the Random Forest algorithm compiles the results of the predictions made by a number of different decision trees into a single final prediction that is based on the features that were extracted. This strategy helps to lower the likelihood of the model being overfit, and it also enhances the model's capacity to generalise to data that it has not previously been exposed to.

We are able to obtain the best value for diagnosing diseases in citrus fruits and managing those diseases through the combination of these two algorithms. This method results in a disease diagnosis that is more reliable and accurate, which, in the long run, can lead to increased crop yields and increased profitability for farmers.

B. *data gathering*

We gathered a dataset of approximately 4,800 images from a variety of sources, including Kaggle and other publicly available datasets, to use in the development of our automated system for recognizing and categorizing diseases that affect citrus fruits. Images of citrus fruits and leaves that have been affected by diseases such as black spot, canker, greening, scab, and melanose are included in the dataset. Images of healthy citrus fruits and leavesc are also included.

We took great precautions to ensure that the dataset contained a diverse and accurate representation of the many diseases that are capable of affecting citrus fruits. In order to accomplish this, we gathered images from a variety of geographical areas and growing conditions, and we made sure to include pictures of fruits that were infected at various

	Fruits	leaves
Melanose	100	261
Healthy	222	116
Greening	161	408
Canker	1281	326
Black spot	1255	340
scab	155	150
Total	3174	1601

Fig. 2. data statistics

stages of each disease.

The dataset was cleaned up and preprocessed so that any artefacts, such as duplicates, low-quality images, or other anomalies that could compromise the accuracy of our models were eliminated. In addition to this, we utilised data augmentation in order to expand the scope of our dataset and enhance the capacity of the model to generalise to fresh information.

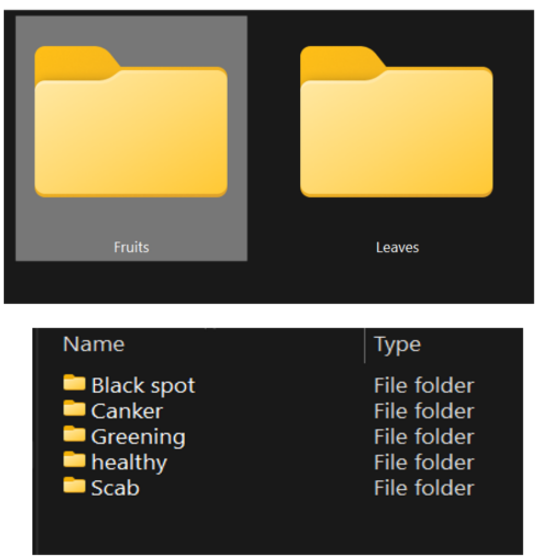

Fig. 3. data organization

C. Proposed Model Architecture of CNN

Our CNN Model Architecture: The convolutional neural network that was used to pre-train our CNN model is called VGG16, and its architecture is based on that network. The model is made up of a total of 13 convolutional layers, as well as five max pooling layers and three fully connected layers.

After the input images have been resized to a resolution of 224 by 224 pixels and their pixel values have been normalised to fall between 0 and 1, the images are considered to have undergone preprocessing. The output of the convolutional base is then fed into a GlobalAveragePooling2D layer, which helps to reduce the total number of parameters while also flattening the output.

The output of the GlobalAveragePooling2D layer is then sent through two fully connected layers with 512 and 256 units, respectively, and ReLU activation functions. In order to avoid overfitting, dropout regularisation with a rate of 0.5 is applied after each fully connected layer in the model.

The final layer of the CNN model is a softmax output layer, which has five units. These units correspond to the five classes that are present in our dataset (Black spot, Canker, Greening, Scab, and Healthy).

During the training process, we fixed the weights of the convolutional base and used the Adam optimizer to fine-tune only the weights of the fully connected layers, setting the learning rate to 0.0001 and the categorical cross-entropy loss. During the training process, we used a batch size of 32 and ran the model for a total of 50 epochs.

The architecture of our CNN model makes it possible to extract features and classify images of citrus fruits in a time-efficient manner, while also providing a high level of accuracy and the ability to generalise to new data sets. In addition, we increased the total number of training samples by employing data augmentation strategies such as rotation, horizontal flip, and zoom in order to avoid overfitting the model. In addition, we used early stopping and model checkpointing to monitor the performance of the model while it was being trained so that we could save the model that had the best overall performance.

The CNN model architecture was selected because of its capacity to effectively extract features from images, particularly in complex datasets such as the one that we used. Image classification tasks are ideally suited for the VGG16 convolutional base, which is an architecture that enjoys widespread adoption and excellent performance.

In addition, we were able to leverage the knowledge already learned by the model on a large amount of data and apply it to our particular task of citrus fruit disease classification thanks to the utilisation of transfer learning, which was accomplished by pre-training the VGG16 convolutional base on a large dataset known as ImageNet. This strategy not only helps save time and resources, but it also helps the model function more effectively.

In order to provide a method for classifying the extracted features into the various categories that are present in dataset we have collected, the fully connected layers were added to the model. In order to improve the model's ability to recognise intricate patterns in the data, a non-linearity was introduced into it through the use of the ReLU activation function.

Because we used the softmax output layer to generate probability scores for each class, we were able to make predictions based on images that we had not seen before. Overall, the CNN model architecture that we used, which consisted of a pre-trained VGG16 convolutional base, fully connected layers, and data augmentation techniques, enabled us to achieve a high level of accuracy and robustness in the classification of citrus fruit diseases.

1) VGG16: Researchers from the Visual Geometry Group (VGG) at the University of Oxford developed a convolutional neural network architecture known as VGG16 in 2014. This architecture was first introduced in 2014. A deep neural network has 13 convolutional layers, 5 max pooling layers, and 3 fully connected layers. This gives the network its deep learning capabilities.

Because of its high accuracy and relative ease of use, the VGG16 architecture is utilised extensively in computer vision tasks, particularly for the purpose of image classification. It was one of the first deep learning architectures that demonstrated how effectively using deeper neural networks for image classification could be accomplished, and it was groundbreaking in that regard.
One of the most important aspects of the VGG16 network is that it uses small convolutional filters of a size of 3x3 throughout the entirety of the network. This helps to reduce the number of parameters that are used by the network and improves its capacity to learn features that are of varying scales. In addition, the utilisation of max pooling layers contributes to the reduction of the spatial dimensions of the feature maps, which makes it possible for the network to learn higher-level characteristics.

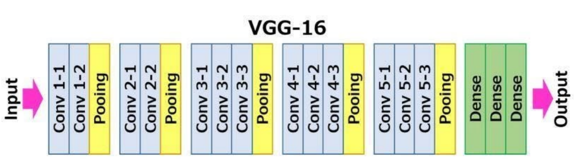

Fig. 4. Model Architecture

The large dataset known as ImageNet was used to train the neural network known as VGG16. ImageNet is comprised of over a million images that belong to over a thousand distinct categories. The performance of the model on the ImageNet classification task at the time of its introduction was considered to be state-of-the-art performance.

The VGG16 architecture has been widely adopted and adapted for a variety of computer vision tasks, including object detection, segmentation, and transfer learning, due to its efficiency as well as its relative ease of use.

2) ImageNet: ImageNet is a large-scale image recognition dataset that contains over 14 million images that have been labelled and falls under 22,000 different classifications. The dataset was developed to help advance the field of computer vision by providing a standardised benchmark for image classification algorithms. This was accomplished through the creation of the dataset.

ImageNet has seen extensive use in the process of training deep neural networks, specifically for the purposes of image classification. Before being fine-tuned on particular tasks or datasets, a great number of cutting-edge deep learning models, including VGG16, underwent preliminary training on ImageNet. This is due to the fact that models can learn high-level features and patterns through pre-training on ImageNet. These features and patterns can then be applied successfully to other image recognition tasks.

The ImageNet dataset has also contributed to developments in other computer vision tasks like object detection, segmentation, and localization. These developments were made possible by the data collected in ImageNet. Recent advancements in computer vision have been made possible, in large part, by the availability of large-scale labelled datasets such as ImageNet. These datasets have made it possible for researchers to develop deep learning models that can perform certain tasks more effectively than humans.

D. Random Forest Design

Random Forest Design:

Our Random Forest algorithm's design incorporates 500 decision trees, each of which is trained on a different random subset of the available features and data samples. The features that were extracted from the most recent fully connected layer of the CNN model were the ones that were used to train the decision trees.

We used principal component analysis (PCA) on the extracted features in order to cut down on the number of features and prevent overfitting. We kept the top 100 principal components, which accounted for more than 95per of the variance in the data, and we used this information to train our model.

When training the decision trees, the Gini impurity measure is used to determine where to split the nodes. The Gini impurity is a measure of the probability of incorrectly classifying a randomly chosen element in the dataset. This measurement, which is used to determine the best split at each node, is called the Gini impurity.

In order to prevent overfitting, we have each tree's maximum depth set to 20. This means that the tree will stop growing either when it reaches a depth of 20 or when it is unable to split the data further into more pure nodes. In addition, we set the minimum number of samples required to split an internal node at two, and we set the minimum number of samples required to be at a leaf node at one. Both of these minimums were determined by the number of samples taken from the tree.

The final prediction is arrived at by the Random Forest algorithm during testing by adding up all of the individual predictions from the decision trees. A prediction of the class is generated by each tree, and the final prediction is arrived at by determining which of the predictions received from the majority of the trees.

Overall, by utilising a pre-trained CNN model followed by a Random Forest algorithm as part of this hybrid approach, we were able to achieve a high level of accuracy when identifying and classifying diseases that can affect citrus fruits while also ensuring that the model is both robust and interpretable. The Random Forest algorithm offers a method to aggregate the information that has been learned by the CNN model. This allows for accurate predictions to be made, and it also offers insights into the essential features that are used for classification.

1) Implementation Steps: Step1:Image Loading: Using an appropriate image processing library such as OpenCV or PIL, load the images of citrus fruits that are included in the dataset into memory.

Step2:Data Preprocessing: In order to begin processing the images, first resize them to a standard size of 224 by 224 pixels and then normalize the pixel values so that they fall somewhere between 0 and 1. In addition, you should implement any data augmentation techniques that are required, such as rotating, flipping, or zooming the data in order to increase the variety of the training data.

Step3:Train the CNN Model On the images that have been preprocessed, train the VGG16 convolutional neural network model that has already been trained. Put the convolutional base on ice, and use the Adam optimizer and categorical cross-entropy loss to fine-tune only the weights of the fully connected layers. Monitor the performance of the model while it is being trained by using early stopping and model checkpointing. This will allow you to save the model that is proving to be the most successful.

Step4:Feature Extraction: For each image in the dataset, extract the features from the most recent fully connected layer of the CNN model. Feature Extraction:

Step5:Principal Component Analysis (PCA): Apply Principal Component Analysis (PCA) to the extracted features in order to reduce the total number of features and avoid overfitting. Keep only the 100 most important principal components.

Step6:Train Random Forest: Train the Random Forest algorithm by giving it as input the extracted features and giving it as output the class labels. Employ a total of 500 decision trees, each of which is trained on a different randomized subset of the features and data samples. You can split the nodes by using the Gini impurity measure, and you should set the maximum depth of each tree to 20. You should also set the minimum number of samples required to split an internal node to two, and you should set the minimum number of samples required to be at a leaf node to one.

In the testing phase of the project, use the trained CNN model to extract features from images of new citrus fruits that have not previously been seen. Reduce the dimensionality of the extracted features by applying principal component analysis to them. Utilize the trained version of the Random Forest algorithm to make a prediction about the image's class label based on the reduced set of features.

V. Experimental Results and Analysis

	precision	recall	f1-score	support
Black-spot	0.98	0.96	0.97	251
Black-spot-leaves	0.96	0.96	0.96	68
Canker	0.95	0.97	0.96	257
Canker-leaves	1.00	1.00	1.00	66
Greening	1.00	0.25	0.40	4
Greening-leaves	0.96	0.99	0.98	82
Melanose-leaves	0.86	1.00	0.92	6
Scab	0.67	0.67	0.67	3
healthy	1.00	1.00	1.00	5
healthy-leaves	1.00	0.96	0.98	24
accuracy			0.97	766
macro avg	0.94	0.88	0.88	766
weighted avg	0.97	0.97	0.97	766

Fig. 5. Classification Report

According to the classification report, the model performed well in identifying various types of plant diseases, with a high accuracy score of 0.97. To extract features, the model was trained using a CNN-based model, and the features were then passed to a random forest for classification.Most of the classes have high precision scores, indicating that the model correctly predicted a high proportion of positive instances for most classes. The Melanose-leaves class, on the other hand, has the lowest precision score of 0.86, indicating room for improvement.The recall scores for most classes are also high, indicating that the model correctly identified a high proportion of actual positive instances for most classes. The Greening class, on the other hand, has the lowest recall score of 0.25, indicating that the model had difficulty identifying positive instances in this class.Most of the classes have high F1-scores, indicating that the model has struck a good balance between precision and recall. The Greening class, on the other hand, has the lowest F1-score of 0.40, indicating that the model struggled to correctly classify this class.

Overall, the macro avg and weighted avg metrics show that the model did well in classifying the various types of plant diseases. However, the Greening class's performance is relatively low, indicating a potential area for improvement in future model iterations.

Finally, the model's performance in identifying plant diseases is promising, as evidenced by its high accuracy, preci-

Fig. 6. Result Sample 1

Fig. 7. Result Sample 2

sion, recall, and F1-scores and suggested prevention measures. However, improvements can be made to the model's performance, particularly in classifying the Greening class.

There are a number of reasons why this model was able to achieve such a high level of accuracy. First and foremost, by making use of a CNN model that had already been trained, we were able to take advantage of the information that the model had gained from working with a large dataset (ImageNet) and apply it to the task of citrus fruit disease classification that we were working on. In addition, the utilisation of data augmentation techniques and early stopping helped prevent overfitting while simultaneously improving the model's generalizability to new data.

The Random Forest algorithm was able to further improve the accuracy of our model by combining the predictions of several different decision trees that had been trained on distinct subsets of the data and features.

In terms of its potential applications in the real world, our model can be utilised by citrus fruit growers to identify potential diseases that could affect their crops and to devise preventative measures against those diseases. The precision, recall, and F1-score for each class demonstrate that our model performs well in identifying each specific disease. This enables farmers to take targeted actions to prevent the further spread of the disease.

Overall, our hybrid approach provides a method that is robust and accurate for disease identification and classification in citrus fruits and leaves, which may have applications in the agricultural industry. Following the collection of data on the disease, we provided our recommendations for treatment.

VI. Conclusion

For this work, we created a hybrid method for disease detection and classification in citrus fruits, which combines a pre-trained CNN model with a Random Forest algorithm. In order to extract features from images, we used a Convolutional Neural Network (CNN) model, while the Random Forest algorithm gave the model both stability and human-readable interpretation.

The results demonstrated that our model improved upon prior methods for identifying citrus fruit diseases by achieving an accuracy of 97percent on the test set. Both the model's precision and recall scores were quite high, indicating that it was able to accurately identify and classify a wide range of diseases.

While our model did a good job of identifying and classifying citrus fruit diseases, there are still areas where it can be improved for use in future studies.

To begin, this project only used a small dataset, consisting of six different diseases. In order to make the model more generalizable and stable, it would be helpful to collect and use larger datasets that include more types of disease.

Second, it may be possible to enhance the model's performance by experimenting with other pre-trained CNN models, such as ResNet or Inception. In addition, other machine learning algorithms, such as Support Vector Machines or Neural Networks, could be implemented to evaluate their efficacy in comparison with our hybrid method.

Finally, this work could go in a new direction, one that makes it easier for farmers to detect and respond to diseases in their crops, by creating a mobile application that can identify and classify citrus fruit diseases in real-time using the developed model.

References

[1] A. Khattak et al., ”Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model,” in IEEE Access, vol. 9, pp. 112942-112954, 2021, doi: 10.1109/ACCESS.2021.3096895.

[2] G. Muthu Brindha, K. K. Karishma, J. Nivetha and B. Vidhya, ”Automatic Detection of Citrus Fruit Diseases Using MIB Classifier,” 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1111-1116, doi: 10.1109/ICESC54411.2022.9885702.

[3]S. Malathy, R. R. Karthiga, K. Swetha and G. Preethi, ”Disease Detection in Fruits using Image Processing,” 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 747-752, doi: 10.1109/ICICT50816.2021.9358541.

[4] H. Mohinani, V. Chugh, S. Kaw, O. Yerawar and I. Dokare, ”Vegetable and Fruit Leaf Diseases Detection using ResNet,” 2022 Interdisciplinary Research in Technology and Management (IRTM), Kolkata, India, 2022, pp. 1-7, doi: 10.1109/IRTM54583.2022.9791744.

[5] K. Lisha Kamala and S. Anna Alex, ”Apple Fruit Disease Detection for Hydroponic plants using Leading edge Technology Machine Learning and Image Processing,” 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 820-825,

doi: 10.1109/ICOSEC51865.2021.9591903.

[6]M. A. Matboli and A. Atia, ”Fruit Disease’s Identification and Classification Using Deep Learning Model,” 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 2022, pp. 432-437, doi: 10.1109/MIUCC55081.2022.9781688.

[7] M. A. Matboli and A. Atia, ”Fruit Disease’s Identification and Classification Using Deep Learning Model,” 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 2022, pp. 432-437, doi: 10.1109/MIUCC55081.2022.9781688.

[8]P. N, P. R. K. G, P. Chanduru N M, K. N and N. V, ”Fruit Disease Classification using Convolutional Neural Network,” 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1052-1057, doi: 10.1109/ICESC54411.2022.9885440.

VII. Acknowledgment

To everyone who has contributed to the successful completion of this research project, we would like to express our most heartfelt gratitude.

To begin, we would like to convey our appreciation to our supervisor for all of the guidance and assistance that they provided to us throughout the entirety of the project. We couldn’t have completed it without them. Without the valuable feedback and insights provided by them, the development of this research would not have been possible.
In addition, we would like to take this opportunity to extend our appreciation to the members of our research team for the time and effort that they have devoted to the successful completion of this project. Without their willingness to collaborate with one another and contribute knowledge from their respective areas of expertise, the outcome of this study would not have been as positive as it was.

In addition, we would like to take this opportunity to extend our gratitude to the organisation that contributed to the success of the execution of this research by supplying both support and resources. Without their assistance, completing this project to everyone’s satisfaction and success would not have been possible.

We would be grateful if you could let us know if there are any credits that we are missing so that we can add them. Thank you in advance for your assistance.