Apple Disease Classifier

HEALTHY

APPLE COD

POWDERY MILDEW

FLYSPECK

W. A. I. M. Weerasinghe
SC/2019/11126

G.D. Asanka Madhushan
SC/2019/11144

V.M. Sooriyaarchchi
SC/2019/11150

R.P.M.L. Vijayathunga
SC/2019/11117

W.K.D.H. Hashan Piyumal
SC/2019/11132

1. Abstract

Fruit diseases are the most crucial problem in economic losses and production in the agricultural industry worldwide. The apple fruit production industry also has been caused by this problem. This paper proposes an image processing approach for identifying apple fruit diseases. According to the current situation, the farmers take the field officers' disease identification and treatment details. But it takes a few days. So, this proposed system can be used to identify apple fruit diseases quickly and automatically. Healthy and three types of apple fruit diseases, namely apple cod, flyspeck, and powdery mildew, were identified using image processing technique in this approach.

Matlab was used as the main implementing software. This proposed approach is composed of the following main steps. Image Acquisition, Image Preprocessing, Image Segmentation, and Classification. Gray-level slicing, RGB channel separation Robert's edge detection, and defected area ratio calculation were used in image segmentation. Before the segmentation, images were preprocessed to predefined resolution using nearest-neighbor interpolation. Then identify and detect each disease using defected area ratio value. Simple mathematic calculation and mid-level image processing techniques learned in the classroom were used in this approach. So we hope to improve this application accuracy furthermore using machine learning in a very practical manner.

2. Introduction

2.1. Problem

Apple fruit is a very popular fruit in all the countries. Because of the nutrients and the health benefits such as helping lower Blood Cholesterol levels, helping for weight loss, preventing tissue damage to Beta cells in the pancreas, and preventing cancer. It has a huge demand in all the countries. Apple fruit is cultivated in many countries like China, the USA, Turkey, Poland, India, Italy, Iran, Russia. In these countries, China is the largest apple producer in the world. World annual apple fruit production is about 87 million MT. In this production, China produces 47 million MT per year. This mass production of apples has increased the diseases and a lot of inconvenience in agriculture. There are several common diseases such as apple scab, black rot, powdery mildew, rust, flyspeck, and apple cod when turning into disease. Some of these diseases infect the tree also. From these diseases, apple cod, flyspeck, and powdery mildew are the most common diseases that infect the apple fruit.

Apple cod (codling moth damage) - Codling moth is the most important insect pest of apple and pears. Damage is done by the larvae, which are cream-colored caterpillars that tunnel fruit and produce ‘wormy’ apples. After the damage, the infected spot of the surface turns dark black.

Powdery mildew - Early infections result in net-like russeting (cork cells) of fruit surfaces as apples mature. Infected fruit may also become distorted and/or dwarfed. Powdery mildew reduces both apple yield and quality.

Flyspeck - Groups of six to 50 or more black and shiny round dots that resemble fly excreta appear on the surface of the fruit. The individual “flyspecks” are clearly separated and can be easily distinguished.
Nowadays technology plays a vital role in all the fields but till today traditional methodologies are used for agriculture. Large-scale agricultural countries like china use technologies like MRI, X-ray imaging for detecting the quality of the fruits. But these technologies are costly for farmers to afford, they occupy large space, users need to have the knowledge to use and analyze the results. Because of these issues, most of the farmers tend to do fruit disease identification manually. In this case, they need experienced people but due to so many environmental changes and lack of resources for getting information, the prediction is becoming tough. So the main purpose of this research is to develop the classifier model using image processing, which will be able to identify apple fruit diseases accurately. Abstract view of the application is illustrated using the below figure.

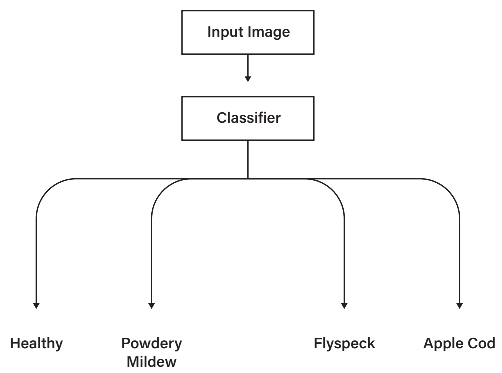

Figure 2.1
Process flowchart of the Disease classifier

2.1. Used Tools

Matlab : Used to implement the application
Photoshop : Used for some preprocessing
Python : Used to generate graphs
Illustrator : Used to design Illustrations, Charts

3. Methodology

There are four phases in this methodology. Those are Image Acquisition, Image Preprocessing, Image Segmentation & Feature Extraction, and Classification.

3.1. Image acquisition

In this phase, all the sample images were collected. We chose both Green and Red apple species in this test. And, amongst the sample images, we included at least five images from each of the 3 Diseases and from healthy apples. One of the major problems was that we could not collect Images on our own due to the pandemic situation, So we had to rely on images that were found on the Internet.

Images of the diseases and healthy apples are shown by figure 3.1, figure 3.2, figure 3.3 and figure 3.4.

Figure 3.1
Healthy Apple

Figure 3.2
Flyspeck Disease

Figure 3.3
Apple Cod Disease

Figure 3.4
Powdery Mildew

3.2. Image Preprocessing

Images that were collected during the Image acquisition phase had unnecessary backgrounds. So first, those backgrounds had to be removed manually. After that, Images were stored on a folder named “input_images” in JPG format, and each was named according to their associated disease. The first two letters were used to denote the disease in shortened form. Next, an arbitrary number was used. Acronyms that were used to denote the disease are mentioned below.

fs : Flyspeck Disease
ac : Apple Cod Disease
pm : Powdery Mildew

ex: “fs-1.jpg”
Means the image contains apple with Flyspeck disease.

When an image is selected as an input, first, It will be rescaled down to 250x250 resolution using Nearest neighbor interpolation. Because using this technique will not reduce smaller features on the images, such as the smaller dots on the Apple caused by Flyspeck Disease.

3.2. Image Segmentation & Feature Extraction

The third phase of the process is Image Segmentation. First, an input image is divided into the Four RGB (Red, Green, Blue) channel matrices separately. The following images show those separated channels.

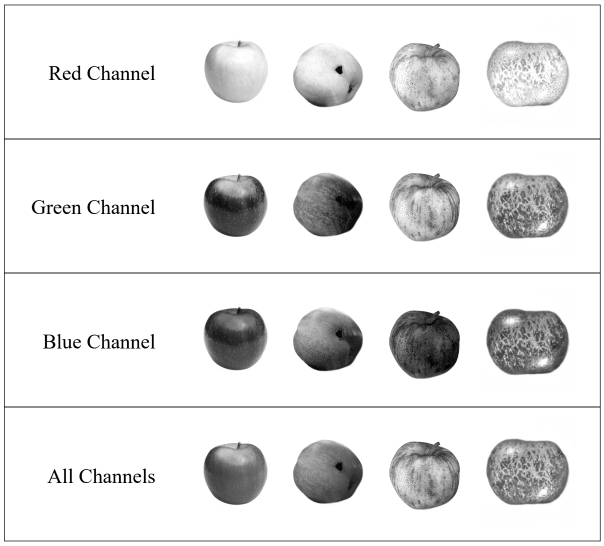

Figure 3.5
Seperation of RGB channels of Healthy, Apple Cod, Flyspeck and Powdery mildew apples respectively.

3.2.1 Classification between Black diseases and Powdery Mildew and Healthy apples

After Channel separation, To segment, the Whole pixels that is of the apple, the Green channel of the Input Image is used to Grey-Level slice intensity values between 250 and 255. This will detect light color pixels on the background and some pixels on the apple if there is any. Then a Binary Image is generated, Pixels that is between 250 and 255 values will be Black on the binary image, and all other pixels will be White. If there were any holes (due to light color pixels on the apple), they would be filled.

Then Red Channel is used to detect darker spots on the apple, So in here also, Grey-Level slicing is done by using Intensity values between 0 and 90. This will also result in Binary Image.

Then, using those two Binary images, the ratio between the Darker spots in the apple vs. the Full area of the apple is taken. The following plot shows the ratio values with their respective disease. From here on, “Black Diseases” will be used to denote Flyspeck and Apple Cod disease as a whole. After running tests on this, a value was chosen to separate between the two. (More on that will be discussed in the Results section)

The following illustrated image shows the whole process mentioned above.

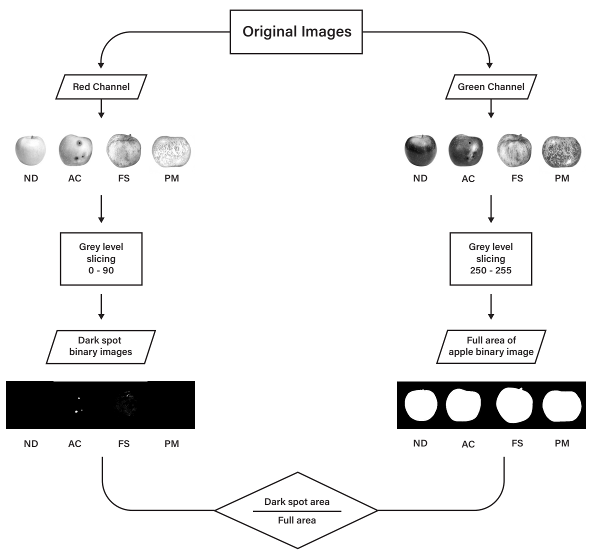

Figure 3.6
Process flowchart of classification between Black Diseases vs Healthy and Powdery Mildew diseases

3.2.2 Classification between Healthy and Powdery Mildew infected apples

Next will be the separation between healthy apples and Powdery Mildew disease. To do this, the Blue Channel greyscale image is used. Here also Edge detection using Roberts edge detection technique is used to identify sudden intensity value changes because, on Powdery mildew diseased apples, powdery like veins cover the apple. Since the powdery-like substance is White like color, an edge detection technique can be used. This operation will generate a Binary Image. As previously, this will also be subject to region separation, but neighboring regions with 8-m connectivity will be considered as a single region. This is due to the continuous vein-like structure of the powdery mildew disease. Also, on healthy apples, there are no sudden color changes, So the edge detection will only detect the border of the whole apple.

The following illustrated image shows the above whole process.

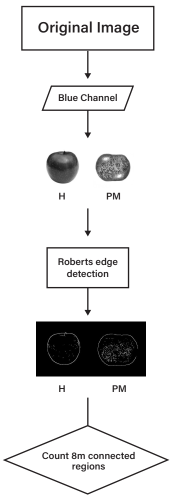

Figure 3.7
Process flowchart of classification between Healthy and Powdery Mildew infected apples

3.2.2 Classification between Black Diseases

Then two operations are done to the Red Channel greyscale image of the input image, they are,

1. Grey-Level slicing (between 0-100)
2. Edge Detection (using Roberts method)

Using the data aquired through these two operations, the separation of Flyspeck Disease and the Apple Cod disease is done.

A binary image is generated after the Grey-Level slicing operation. In this image, Dark spots that were identified in the previous Grey-Level slicing will be included, along with some slightly lighter dark spots. That means Intensity values between 90 and 100 will also be included. This is because some smaller dark spots occurred due to Flyspeck disease can have less dark spots, so intensity values between 0-100 will detect those also.

Another Binary image is also generated by using the Roberts Edge Detection technique. This is used to identify sudden color variations on the image, So Borders of darker spots and the border of the apple will be highlighted. Also, Continuos dark spots (like the ones in Apple Cod disease) will not be highlighted; only their borders will be.

So by using those two Binary Images, a new Binary Image will be generated that will contain pixels that are common to the old two binary images. Those common pixels are mostly the ones that were generated due to small dark specks caused by Flyspeck disease. This can be used to differentiate between the two black diseases. This new Binary Image will be constructed using Binary Arithmetic AND operation.

Then the Binary image is divided into regions, in here, neighboring regions with 4-m connectivity will be counted as separate regions. This is done to further remove regions that might have been generated due to shadows or dark spots that were generated due to Apple Cod disease. After that, the number of regions will be used to differentiate between Apple Cod and Flyspeck diseases.

The following diagram illustrates the above mentioned process.

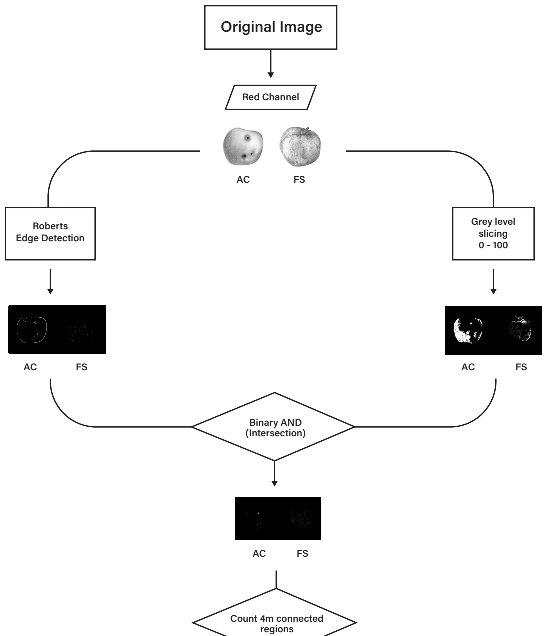

Figure 3.8
Process flowchart of classification between Apple Cod disease and Flyspeck Disease

3.2. Classification

Previously mentioned main three processes take a considerable amount of time to classify between all three apple disease types and healthy ones if it was done in a step-by-step way. Most of that time will be consumed by all three Grey-Level slicing operations. So, to reduce the time it takes, those three main processes are done parallelly.

After all the data for an input image has been collected, the classification is done using the previously obtained values, which are,

- 0.0067 - To seperate between Black vs healthy and Powdery mildew disease.
- 47.5 - To seperate between Flyspeck vs Apple Cod disease.
- 280 - To seperate between Healthy vs Powdery Mildew disease

How these values were obtained will be discussed in the next chapter

3. Results

3.2.1 Classification between Black diseases and Powdery Mildew and Healthy apples

The following plot shows the collected data (ratio between Dark area and the Full apple area) for all sample input images.

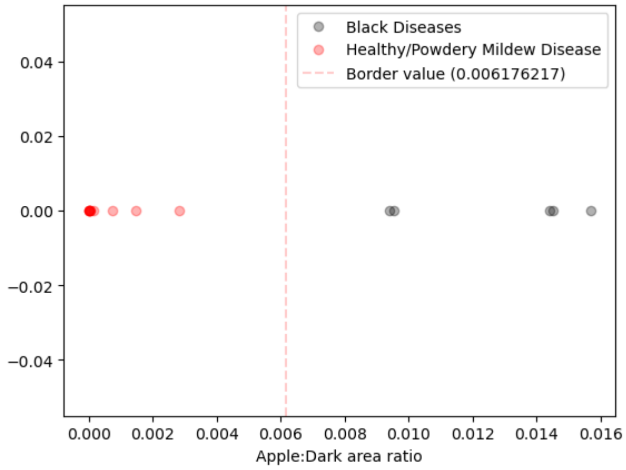

Figure 3.5
Ratio values of Apple area to Dark spot area on the apple.

Then using the data above, a suitable value was chosen by taking the midpoint between the lowest value from Black diseases and the Highest obtained value from the Healthy/Powdery Mildew infected apples to differentiate between diseases that cause dark spots (Flyspeck and Apple Cod) on the apple vs. Healthy apples and Powdery Mildew diseases. And that value is ***0.006176217***.

3.2.2 Classification between Black Diseases

The following plot shows the data (number of regions) obtained for all Apples that were infected by Flyspeck or Apple cod disease.

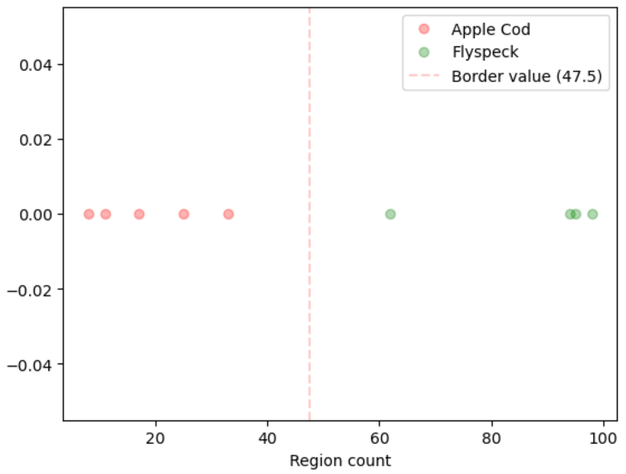

Figure 3.6
Dark Region count on Flyspeck and Apple Cod diseases

Using that data, we can determine a value to differentiate between the two by taking the midpoint between lowed recorded Flyspeck data point, and Highest recorded Apple Cod data value. And that value is ***47.5***

3.2.2 Classification between Healthy and Powdery Mildew infected apples

The following plot shows the collected data (number of regions) for all sample Healthy and Powdery Mildew infected apples.

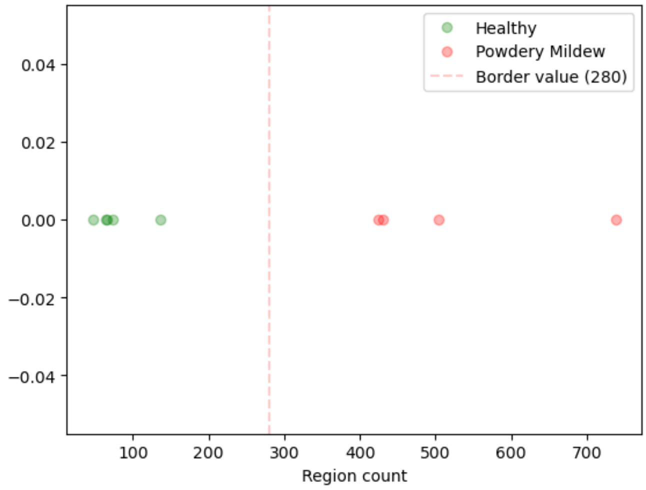

Figure 3.7
Sudden intensity change regions on Healthy and Powdery Mildew infected apples

Using the same method as previous cases, the midpoint between highest and lowest data points of the Healthy and Infected apples were used to get the value to differentiate between the two. And that value is ***280***.

5. Discussion

5.1. Importance & Advantages

- **Low Cost & Resource allocation**

Large-scale agricultural countries like china use technologies like MRI, X-ray imaging for detecting the quality of the fruits. But these technologies are costly for farmers to afford, they occupy large space, users need to have the knowledge to use and analyze the results. Because of these issues, most farmers tend to make fruit disease identification manually. In this case, they need experienced people, but due to so many environmental changes and lack of resources for getting information, So this tool can be utilized with simple machinery to classify between diseased and healthy apples.

- **Minimize damage to crops due to diseases**

Some infectious diseases can cause a lot of damage to both crops and plants. So early identification of those highly infectious diseases can help farmers to eliminate them and potentially save crops and minimize the damage. This tool can be used to identify between those diseases and give early warning to the users if it was able to find any infectious diseases.

- **Further classification of Healthier Apples**

This tool can be further developed to classify between Healthier apples to find apples with much higher quality without any effort from farmers, So farmers can easily sell them at a higher price.

5.2. Drawbacks & Limitations

- **Shadows & Highlights**

Input images that contain Shadows or Highlighted areas due to flashlights and environmental light conditions could output wrong results.

Solution 1 : When this application is utilized with machinery, those machineries can be designed to take pictures of apples with correct light conditions to prevent them.

Solution 2 : Further develop the application to detect shadows and other light conditions and avoid them.

- **Segmentation of the Apple area**

A major drawback was that the sample input images had to be preprocessed manually to remove unnecessary background and replace them with a white background. The application heavily relies on the background to identify the whole apple.

Solution 1 : Implement Machine Learning techniques to identify the apple automatically without human input.

Solution 2 : Input images can be taken on a white background or any other background that can be separately identified with the apple. So, machinery can be designed specifically to take pictures of the apples with a certain background (ex: White Background).

- **Lack of data**

Results that were obtained from Input images are only from 20 images, with 5 per disease/healthy.

Solution 1 : Collect more data by taking pictures of apples and further improve the model.

5.3. Gained Knowledge

- **Preprocessing of Images**

One of the major problems that occurred during this project was the preprocessing of the input images. Many tests were conducted with images that are of different resolutions, and this resulted in wrong outputs. So, the Images were downscaled to a specific size. But that leads to other problems like reduction of features on the images.
For example, If an Image with Flyspeck disease was rescaled to a smaller size with Bicubic interpolation, this results in reduction/fading of small black specks. So a different image resizing technique had to be adopted. So Nearest Neighbor technique was used because this technique preserves small details.

- **RGB Channel seperation**

Another major problem was Feature Extraction. Many techniques were tested to extract unique features for a specific disease to differentiate between them. One of the easiest and an effective technique was separation of color channels.
Channel Separation can be easily adapted to extract features that are unique to a specific color.
For example: Identification of Dark spots on the image can be easily done using the Red Channel and Green Channel. Because apples are usually Red or Green, so with those channels separated, Dark areas can be identified easily.
We hope to test out other color models like,

- **HSV**
- **L*a*b***

5.4. Further Work

- **Implement Machine Learning**

Currently, the values that are used to classify diseases and healthy apples are manually calculated. So we hope to improve this by adopting Machine Learning algorithms with a much larger dataset. This will greatly increase the accuracy of the application.
Also, this tool only works for input images of apples with White background. We hope to improve on this also by adopting Machine Learning so that Apple Objects can be automatically identified and then preprocessed.

- **Implement a mechanical aspect**

This tool can be utilized with mechanical machines to physically separate healthier apples from diseased ones. So simple, low-cost microcontrollers (like Arduino, Raspberry Pi) can be used to implement that.

- **Implement multiprocessing techniques**

We hope to implement multithreading features to the tool to speed up the process even more. So time it takes to classify bulk of input images will be significantly low.

6. References

[1] *R. Gonzalez, R. Woods, Digital Image Processing 3rd ed.*

[2] *Matlab documentation*

[3] *Dr. Nusrath Hameed CSC2143 Lecture Notes*

[4] *Common Apple Diseases*

[5] *Apple Industry*