A Progress Report

on

Lesion Isolation from Infected Plant Leaves using Image Processing and Machine Learning Techniques

carried out as part of the course CSE CS3270 Submitted by

Aryan Gupta

199301088

VI-CSE

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

In

Computer Science & Engineering

MANIPAL UNIVERSITY JAIPUR

INSPIRED BY LIFE

Department of Computer Science and Engineering,
School of Computing and IT,
Manipal University Jaipur,
May 2022

<u>CERTIFICATE</u>

This is to certify that the project entitled "**Lesion Isolation from Infected Plant Leaves using Image Processing and Machine Learning Techniques**" is a bona fide work carried out as part of the course ***Minor Project : CS3270***, under my guidance by ***Aryan Gupta***, student of Bachelor Of Technology (B.Tech.) in Computer Science and Engineering (CSE) at the Department of Computer Science & Engineering , Manipal University Jaipur, during the academic semester *VI of year 2021-22.*

Place:

Date: Signature of the Instructor (s)

<u>DECLARATION</u>

I hereby declare that the project entitled "**Lesion Isolation from Infected Plant Leaves using Image Processing and Machine Learning Techniques**" submitted as part of the partial course requirements for the course Minor project (***CS3270)*** for the award of the degree of Bachelor of Technology in Computer Science and Engineering at Manipal University Jaipur in the semester during academic year 2021-22, has been carried out by me. I declare that the project has not formed the basis for the award of any degree, associate ship, fellowship, or any other similar titles elsewhere.

Further, I declare that I will not share, re-submit, or publish the code, idea, framework and/or any publication that may arise out of this work for academic or profit purposes without obtaining the prior written consent of the Course Faculty Mentor and Course Instructor.

Signature of the Student:

Place:

Date:

ABSTRACT

According to the Food and Agriculture organisation of United Nations, the land area harvested for Apples stood at 308,000 hectares in 2020 making it one of the most popular fruit choices in the country. India's Apple imports more than doubled in 2021 according to the Economic Times. Apple can mostly grow in certain regions of India like Jammu and Kashmir, Himanchal Pradesh, hills of Uttarakhand and Uttar Pradesh.

Apple Black rot is caused by a fungus named Botryosphaeria obtuse. Initial infections on leaves appear as reddish-brown flecks that enlarge to circular brown lesions (4-8 mm diameter). [1] Venkatasubbaiah, P. et al. (1991) The fungus Botryosphaeria obtusa can cause devastating losses to Apples and crab-Apples over a wide geographical range and especially in warm humid climates. [2] Apple Scab is caused by the Venturia inaequalis fungus. Apple scab occurs everywhere in the world where apples are grown and results in more losses than any other apple disease. Apple scab infections are start in spring on emerging and young leaves. Early lesions appear 10 days later as lighter green areas compared to the surrounding leaf tissue. Lesions increase in size and can grow up to 1cm in young leaves. [15]

The lesion isolation or the masking technique developed so far in this project would be extremely useful for segmentation of the Apple black rot disease. Once the ROI is isolated, further processing can be done in the form of classification of the disease, the stage in which the disease is, correlation between the behavior of certain diseases, and so on. This will help in making better choices in chemicals, insecticides, fertilizers and agricultural practices in general which will minimize disease, maximize yield and skyrocket profit.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

According to Food and Agriculture organisation of the United Nations, approximately $220 billion is lost every year due to plant diseases. Outbreak of such diseases, if not controlled, has been leading to massive famines with people having to either starve to death or relocate. The Bengal Famine of 1943 was the result of a combination of human disasters and a devastating disease of rice called brown spot caused by Bipolaris oryzae. The Bengal Famine led to the death of an estimated 4 million people in India. [3]

The need for innovation in agricultural practices is rising every day and it is one of those fields often undermined when it comes to use of technology and artificial intelligence which basically present in a lot of aspects of our day-to-day life whereas we're still using decades old practices in agriculture which especially affects our agrarian economy. Currently farmers depend on manual visual means (naked-eye) for investigation of diseases in their crops which doesn't scale well for large farms as it is too expensive and time consuming on check on such huge distances, leaving a dangerous level of risk for a plant-disease outbreak. Therefore, the automated use of image processing and machine learning together can help provide farmers with some deeper insights about their crops and help them make better decisions in their farming.

The isolation of ROI in an infected plant leaf is a really important step when it comes to solving a big problem of curing plant diseases. As the population booms, the demand for food increases. The harvesting, packaging, storing and delivering of the crops can't be made faster after a point so all the focus is on maximizing the yield which leads to poor farming practices like disturbing the soil composition, using excessive fertilizer, monoculture etc.

LITERATURE REVIEW

Review of a few existing techniques –

[4] describes a simple technique for lesion isolation with only colour channel manipulations and Boolean operations on binary masks where there are four binary masks generated and combined into one single segmentation technique.

[5] talks about the use of masking techniques to mask out the green pixels as they don't play much role in lesion isolation and significantly increase the computational time. It then uses principal component extraction to extract relevant features from the leaf. K-means clustering is applied to form clusters of affected areas followed by SVM classification to predict the output.

[6] is about using eigen spectrum modelling followed by eigen feature extraction and regularization algorithm which produces a 90% accuracy of detecting the red spot disease affecting cotton.

[7] proves that after applying k-means segmentation on the leaf, SVM produces a better accuracy than ANN on diseases like Alternaria, Apple black spot and Apple leaf miner pest.

[8] discusses the importance of feature selection using fuzzy curves which significantly reduces the dimensionality of the original feature set and boosts the accuracy of the classifier's performance.

[9] introduces a simple algorithm that changes the histograms of h (from HSV) and a (from LAB) channels to create a powerful technique that produces good results on a plethora of diseases despite of leaf-colour, symptom-colour and illumination variations.

[10] works on sugarcane diseases and compares ANN, linear SVM and non-linear SVM and concludes that non-linear SVM (at 94% accuracy) is the best classifier for that algorithm. During pre-processing, it works on the LAB colour space and applies k-means segmentation, canny edge detection and feature selection which reduces the computational time for the

classifier.

[11] takes one Northern leaf blight affecting corn and applies active contour model, said to be one of the most powerful image segmentation techniques, and oppositional fruit fly algorithm to create a global region-based segmentation which has a precision value of 98.26 and 97.08 for OFA and ACM respectively.

[12] divides the LAB colour space to L and AB percentage-wise in different cases and uses them as layers for the CNN model namely 20%L-80%AB, 50%L-50%AB and 80%L-20%AB respectively. They produce different accuracies in different conditions like blur, motion-blur and salt and pepper (4%) with consistent 99%+ accuracy in different cases.

[13] uses genetic algorithm for segmentation into clusters to produce an average accuracy of 97.6% which yielded consistently better results compared to k-means segmentation with the same classification algorithm in working on diseases of plants like lemon, beans and banana.

Table 1. Review of related works

Paper Name	**Disease Name/Type**	**Pathogen**	**Technique**
[4] “A new automatic method for disease symptom segmentation in digital photographs of plant leaves”	Homogeneous Dark Homogeneous Bright Heterogeneous Powdery White Small Dark Small Bright Colour Changing	Phoma costaricensis Phaeoramularia manihotis Citrus leprosis virus C Mycosphaerella citri Xanthomonas citri	Combination of four binary masks
[5] “OpenCV Based Disease Identification of Mango Leaves”	Gall Midge infestation	Gall midges (plant-parasitic insect)	k-means clustering

[6] "Diseases Classification on Cotton leaves by Advance Digital Image Processing Approach"	Leaf Crumple Red Spot Anthracnose Leaf Roll Crown Gall	Agrobacterium tumefaciens Geminiviridae family Colletotrichum	Eigen Spectrum Modelling
[7] "Potential of radial basis function-based support vector regression for Apple disease detection"	Apple leaf miner pest Apple black spot Alternaria spot	Botryosphaeria obtusa Western tentiform leafminer (insect)	support vector regression
[8] "Features selection of cotton disease leaves image based on fuzzy feature selection techniques"	Anthracnose Black Spot Red Leaf Blight	Pyrenochaeta glycines Colletotrichum	Fuzzy curves
[9] "A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing"	Cercospora leaf spot Powdery mildew Alternaria brown spot Citrus greasy spot Diplodia leaf streak Myrothecium leaf spot	Cercospora abelmoschi Erysiphe, Microsphaera, Phyllactinia, Podosphaera, Sphaerotheca, and Uncinula.	Colour channels histogram manipulations

[10] “Application of Image Processing Techniques in Plant Disease Recognition”	Sugarcane red dot Sugarcane mosaic virus Brown spot	Colletotrichum falcatum Aphids Bipolaris sacchari	K-means segmentation, SVM
[11] “Leaf Disease Segmentation From Agricultural Images via Hybridization of Active Contour Model and OFA”	Northern leaf blight	Exserohilum turcicum	Active Contour Model and OFA Algorithm
[12] “Reliable Deep Learning Plant Leaf Disease Classification Based on Light-Chroma Separated Branches”	Apple Black Rot	Botryosphaeria obtuse	CNN
[13] “Detection of plant leaf diseases using image segmentation and soft computing techniques”	Early scorch disease Frog eye leaf spot	Xylella fastidiosa Cercospora sojina	Genetic algorithm

Problem statement – The extraction or isolation of the region of interest that is, lesions affecting a plant leaf using image processing and machine learning techniques.

Objectives –

1) To develop a method to segment the disease-ridden tissue from a healthy leaf tissue.

2) To use the algorithm to make a more generalized technique that works for more than one disease that is both accurate and robust.

METHODOLOGY AND FRAMEWORK

The dataset used for the project is the PlantVillage Dataset. It provides three types of image data – colour, segmented and grayscale. For our use, we use the segmented images as it helps to focus the isolation only of the leaf surface rather than the complete picture. For the current disease, Apple Black Rot, we have 621 images in all three categories which are uploaded to Google drive and then imported to Google Colab to run our algorithm.

Algorithm –

We classify the Apple black rot segmented dataset into three categories according to the spread – light, moderate and high spreads respectively. The following criteria is used for the classification based on these features –

Light spread – One single big spot on the boundary or one single big spot on the boundary with a couple small spots in the interior or 1-4 independent spots towards the interior.

Moderate spread – More than 4 independent spots towards the interior.

High spread – More than 10 independent spots covering a major portion of the leaf.
This enables us to work on simple images first (light spread) and then move our way up to more complex images.

We used these features based on purely visual references and manual estimations to identify three distinct categories on the entire dataset as it is a highly subjective matter.

We started with the light spread which was relatively easy and working fine for certain threshold values and the R channel of RGB as mostly the entire infection was at one spot so it was easier to separate. But ultimately, this method failed to generalize on other categories where the infection was scattered as the threshold value was different for different spots. We then observed that the H channel of HSV produced distinct spots of Apple black rot visible in black on a greyish leaf. Isolating the R from BGR and A from LAB also showed clear infected spots but they were harder to isolate as there were still varying intensity values all around the leaf which is harder to produce a mask on.

Finally, a binary inverse thresholded image is derived from the original using a threshold value of 30, a value that was tested on all classes and produced quite accurate results. A bitwise AND operation is performed on thresholded image and RGB image to get the isolated spots on the leaf.

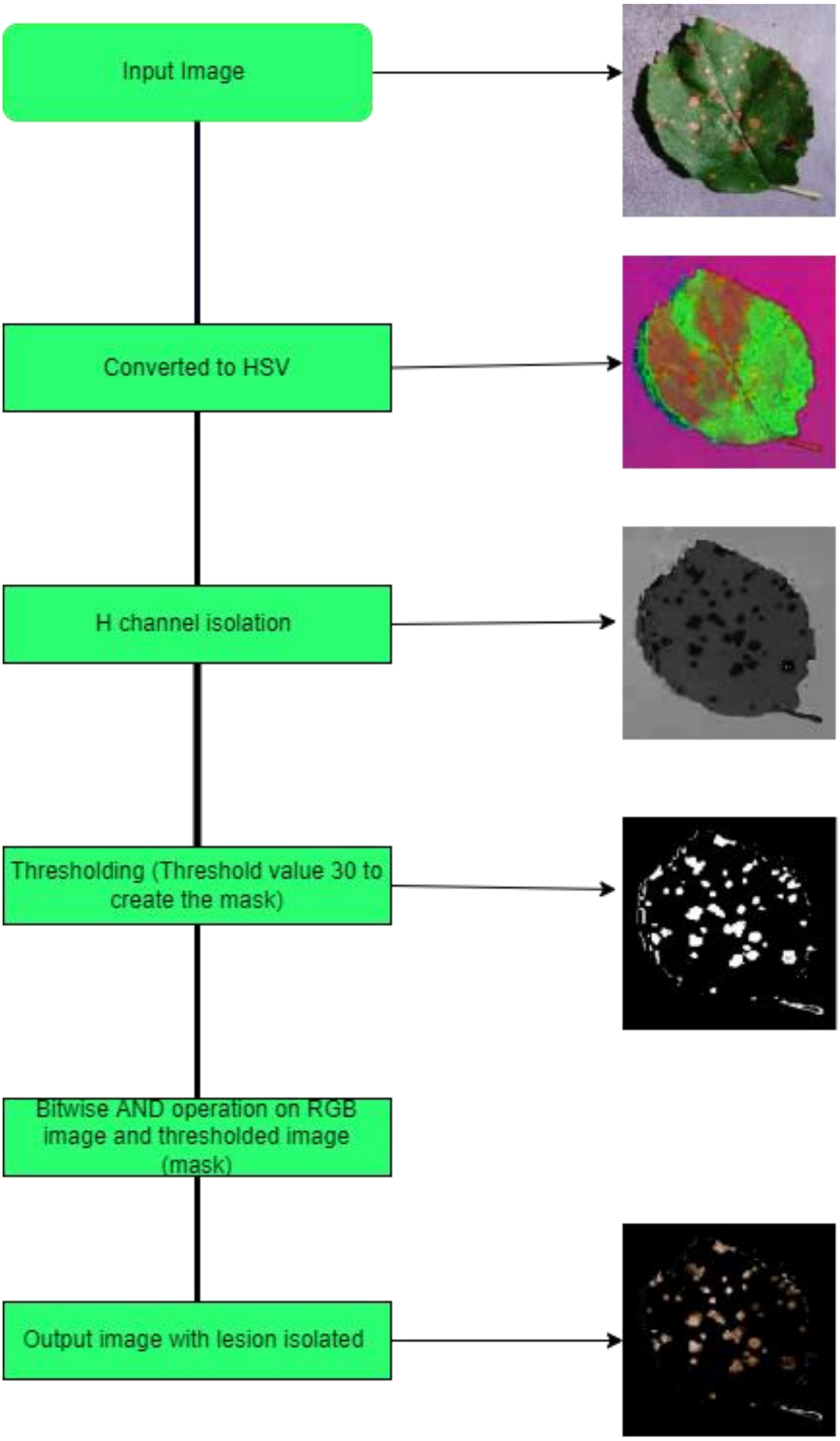


Figure 1. Flowchart explaining algorithm implementation

WORK DONE

Successfully built a simple and robust algorithm for segmentation of lesions of Apple Black Rot on an infected leaf using several image processing techniques implemented by OpenCV libraries.

Table 2. Gantt Chart

		Months								
		January	February		March		April		May	
	Schedule for Project Work (Green - completed)	15-30	1-15	15-30	1-15	15-30	1-15	15-30	1-15	15-30
1	Initial Research									
2	Synopsis submission									
3	Setting up basic requirements									
4	Developing a mask for light spread (Apple black rot)									
5	Developing final mask for Apple black rot									
6	Mid Term Presentation									
7	Solving pre–Mid Term challenges									
8	Creating a universal technique									
9	Project Report									
10	Final Presentation									

The proposed algorithm was implemented on Google Colab in a PC with: Processor – Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz 1.80 GHz, RAM – 8.00, System Type – 64-bit operating system, x64-based processor, Python version - Python 3.6. 9 (on Google Colab)

Table 3. Algorithm output on light, medium and high spreads

High	**Medium**	**Low**
		
		

After we finished our work with the mask for Apple black rot. We moved to Apple scab after several unsuccessful attempts with Apple rust disease as mentioned in the discussion. So we decided to use a software called ImageJ which provided a plethora of tools and plugins for which we were writing a bunch of code for, which made our visualization easier and much faster on the go. We tried quite a few segmentation techniques on Apple scab on both coloured and segmented images. Most of them had problems like in the case of Apple rust except one technique that produced much better results hence a more accurate mask overall. Few of the attempts are discussed below:

1. **RATS thresholding** –

Stands for Robust Automatic Threshold Selection. This technique works well with noisy images and variable background but fails to detect minute changes in intensity like in our case.

Parameters set as:

Noise threshold – 25
Lambda factor – 3
Min leaf size – 51

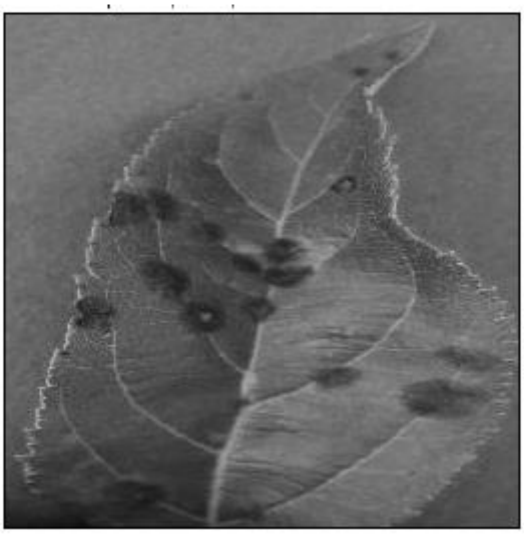


Figure 2. Input image (Green channel of the RGB image)

Figure 3. Output of RATS thresholding

2. **Morphological segmentation** –

Morphological segmentation works on applying morphological operations on the image and auto detecting a certain threshold based on input paramters. We use the watershed segmentation model for our image trying to find the right threshold values but it doesn’t work as intended.

Input parameters –

Input image – border
Tolerance – 30

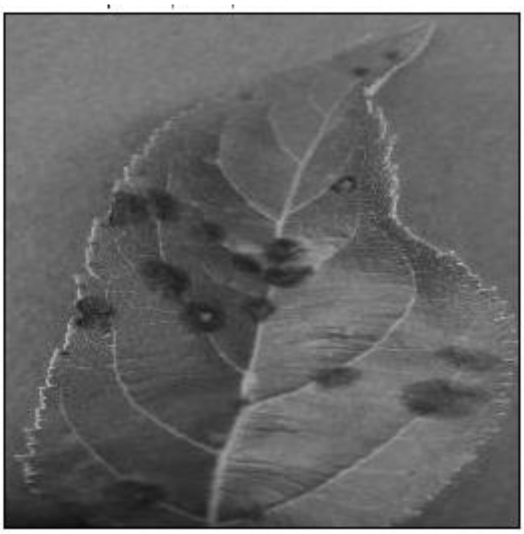


Figure 4. Input image (Green channel of the RGB image)

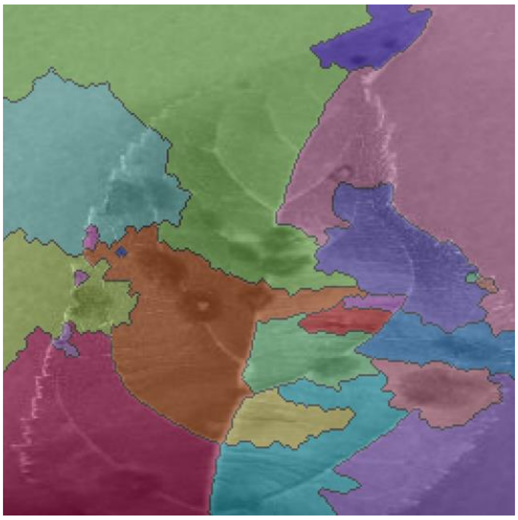


Figure 5. Output after watershed segmentation

3. **Mixture Modelling thresholding –**

With outputs as –

Mu1 : 60.14163, variance1 : 140.61562

Mu2 : 105.830284, variance2 : 246.22539

ErrorMin : 15282.274

Diff Mu : 45.688656

Direct threshold : 75

Real threshold : 74.0

Input being the same green channel image –

We get the following thresholded image

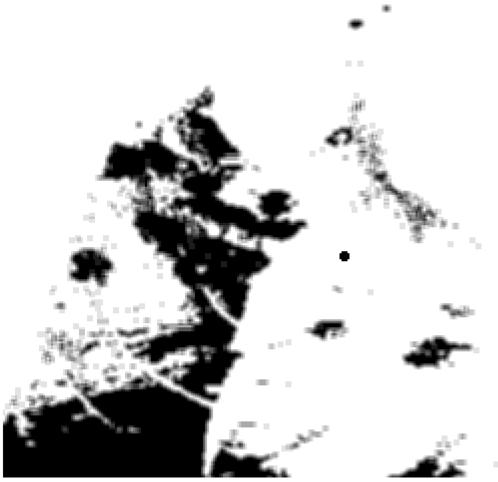


Figure 6. Result of Mixture modelling thresholding

Even **grayscaling and enhancing contrast** to 3.0% showed clear spots of the disease but still not enough compared to the solution techqniue.

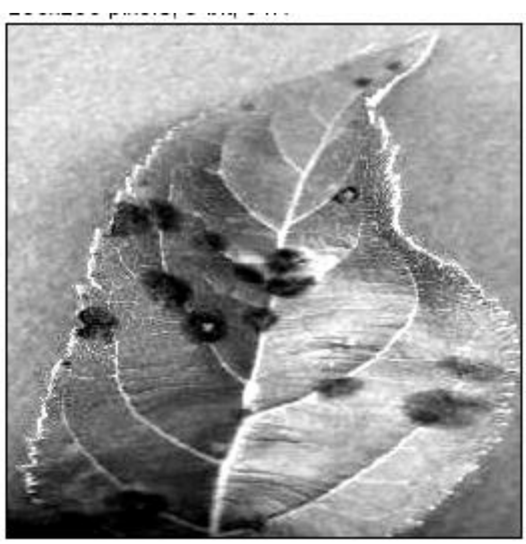


Figure 7. Contrast enhancement result on 8-bit grayscaled image

Another **custom technique** was used as the image was converted to RGB first and then thresholding was done with various threshold values but this was the best result we could obtain on a threshold value of 55 , which proves that the technique which worked for Apple Rot doesn’t work well for us in the case of Apple Scab.

Output image –

Figure 8. Output of the custom technique

Best performing model for Apple scab was a custom mask –

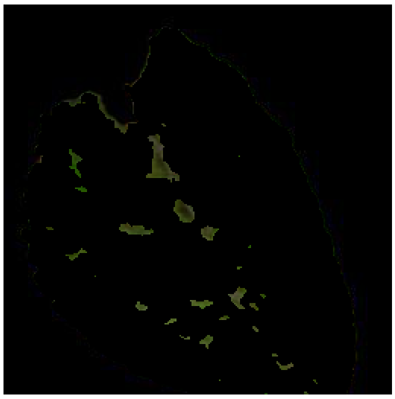
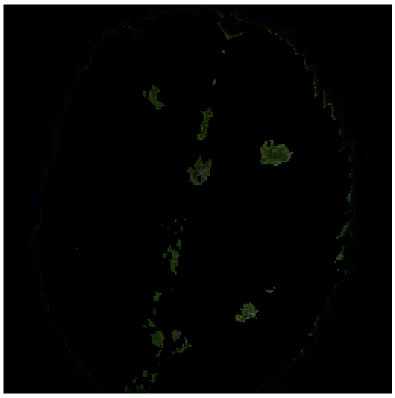
The difference between black rot and scab was that H channel couldn’t produce as distinct or even remotely distinct disease spots on the leaf, whereas green and lightness channels of RBG and LAB respectively had a pretty similar results in terms of a fair amount of intensity separation between the spots and the rest of the leaf but at certain parts they both matched so there was a need to increase the contrast which we did using Contrast Limited Adaptive Histogram Equalization or CLAHE technique on segmented images of the dataset.

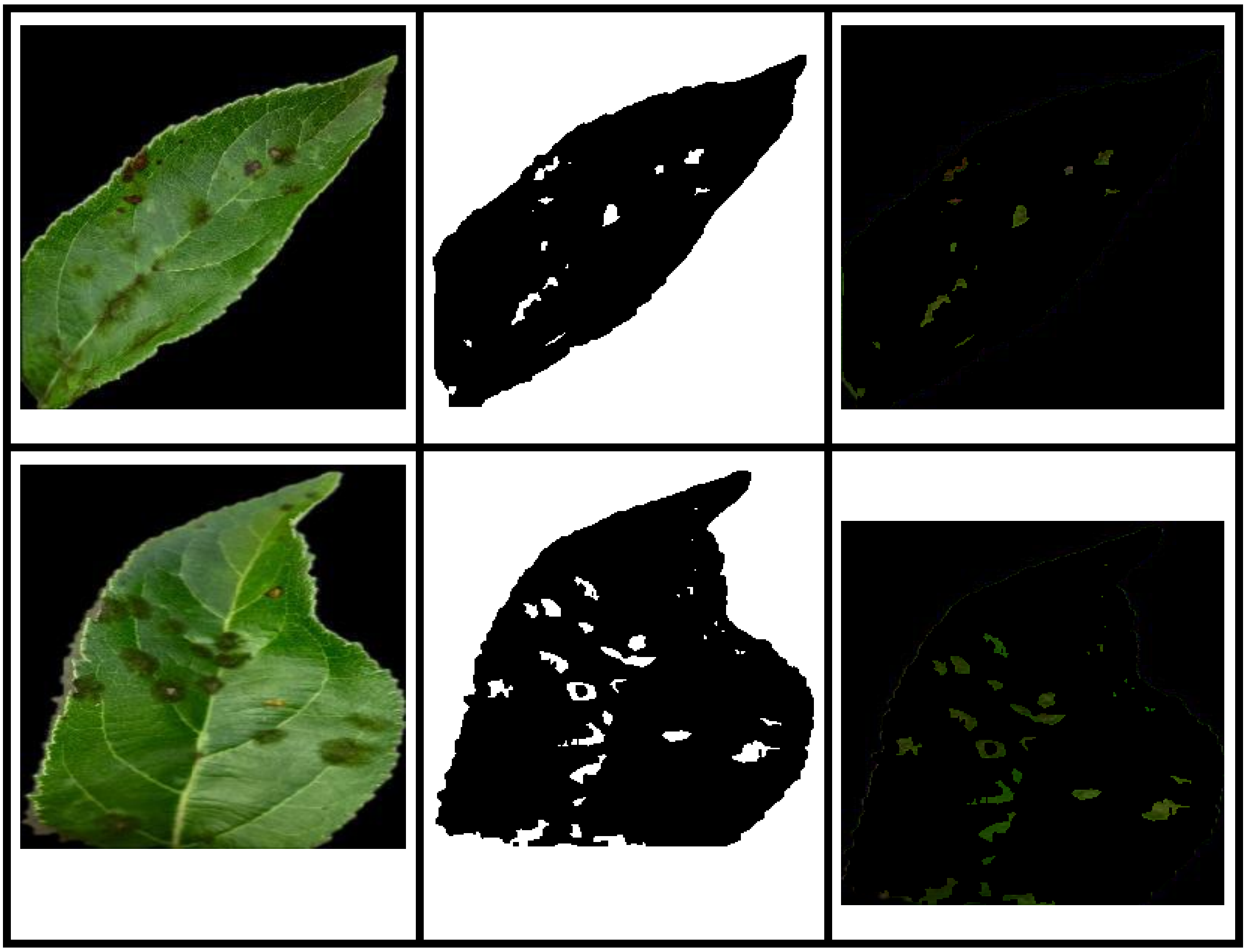
At first the resultant thresholded images were quite grainy with a noticable particle size which interferred with the accuracy of the mask so we used both Gaussian blur and erosion to get rid of that. We applied binary inverse thresholding and set a threshold value of 60 in this case as it generally performed better than other values but we weren’t able to apply universal technique to find a global minima or maxima and then try to find an automated threshold value from the leaf’s intensity distribution as it was varying a lot from leaf to leaf.

The thresholded image and the original image were passed through a bitwise AND operation and we get the isolated lesions from the image which range from brownish to greenish brown in colour which subtle differences.

Results –

Table 4. Custom mask results on Apple Scab disease

Original	**Mask**	**Mask applied**
		
		



Discussion –

One of the difficulties faced was the pixelating of certain images in the HSV channel. There was no specific pattern found, like illumination, or spread or leaf-colour that could explain the error. Earlier it was assumed it is because of the shadow but for a lot of leaves with shadow in the background, the algorithm seemed to work just fine. Tried the technique on Jupyter Notebook assuming the problem was with Google Colab but that failed too.

For light spread another technique that was talked about in the methodology section worked well too, with R channel being isolated and truncated thresholding technique applied. Tried morphological operations like erosion, dilation with different parameters but they did not really affect the accuracy. We haven't made a mask for the leaf petiole yet as it's of a different colour, size, thickness and position in every image.

For the Apple Rust disease mask, there were a number of problems faced – the spots are not as prominent as in Black Rot, therefore it is hard to segment it. Tried making custom masks by hardcoding the range of pixels of the disease area but it fails to scale up to other leaves. The spots range from whitish red to brick/blood/rust red (a highly variable color range) and HSV is not working for them. The spread isn't clear for any color space of any color channel. Using histogram equalization in different color spaces but the lesions aren't separating enough to apply any thresholding technique. In a lot of leaves, there are large patches of discoloration on the leaf surface and mixed with the spots which are difficult to work with.

Individual Contribution –

Aryan Gupta – Did a course on OpenCV and learnt imagej tool operations, went through more than 20 research papers to learn about lesion segmentation techniques to create masking algorithms.

Shaleen Poddar – Did a course on OpenCV, researched about image segmentation to provide valuable inputs.

CONCLUSION AND FUTURE PLAN

The proposed technique is simple and has a low computational cost which will help in scaling it to a multitude of diseases as it only uses color spaces and channels and basic thresholding which are not plant-specific, also we work on the HSV color space is a color independent space model. However, it fails to scale up to the Apple Rust disease as it works well when the spots are clear and more pronounced. So, the future plan would be to develop a technique had addresses the issues with that disease or create a generalized technique for all Apple diseases or all Black Rot diseases.

Proposed workplan of the project –

We plan to create a more accurate mask for Apple scab by overcoming the technical difficulties and researching more about the topic to be more segmentation techniques available on different software or create a custom mask like Apple black rot. When successful, we would be creating universal technique for at least three apple diseases that will be simple, accurate and robust.

REFERENCES

[1] Venkatasubbaiah, P., Sutton, T. B., & Chilton, W. S. (1991). Effect of phytotoxins produced by Botryosphaeria obtusa, the cause of black rot of Apple fruit and frogeye leaf spot. Phytopathology, 81(3), 243-247.

[2] Helga George (2019). "HOW TO IDENTIFY AND CONTROL APPLE BLACK ROT AND FROGEYE LEAF SPOT" https://gardenerspath.com/how-to/disease-and-pests/Apple-black-rot-frogeye-leaf-spot/ (April 5, 2022)

[3] Jeff Batten (1999). "World Food Crisis: Meeting the Demands of a Growing Population" https://www.apsnet.org/edcenter/apsnetfeatures/Pages/FoodCrisis.aspx (April 5, 2022)

[4] Barbedo, Jayme. (2017). A new automatic method for disease symptom segmentation in digital photographs of plant leaves. European Journal of Plant Pathology. 147. 349-364. 10.1007/s10658-016-1007-6.

[5] Sethupathy, Jayaprakash & Veni, S.. (2016). OpenCV Based Disease Identification of Mango Leaves. International Journal of Engineering and Technology. 8. 1990-1998. 10.21817/ijet/2016/v8i5/160805417.

[6] Gulhane, V.A., & Gurjar, A. (2012). Diseases Classification on Cotton leaves by Advance Digital Image Processing Approach.

[7] Omrani, Elham & Mohtasebi, Saeed & Band, Shahab & Saboohi, Hadi & Anuar, Nor & Md Nasir, Mohd & Khoshnevisan, Benyamin. (2014). Potential of radial basis function-based support vector regression for apple disease detection. Measurement. 55. 10.1016/j.measurement.2014.05.033.

[8] Yan-Cheng Zhang, Han-Ping Mao, Bo Hu and Ming-Xi Li, "Features selection of cotton disease leaves image based on fuzzy feature selection techniques," 2007 International Conference on Wavelet Analysis and Pattern Recognition, 2007, pp. 124-129, doi: 10.1109/ICWAPR.2007.4420649.

[9] Barbedo, J. G. A. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. Tropical Plant Pathology, 41(4), 210–224. https://doi.org/10.1007/S40858-016-0090-8

[10] Renugambal. K, Senthilraja. B, 2015, Application of Image Processing Techniques in Plant Disease Recognition, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 04, Issue 03 (March 2015), http://dx.doi.org/10.17577/IJERTV4IS030829

[11] Jayanthi, M. & Shashikumar, D. (2020). Leaf Disease Segmentation From Agricultural Images via Hybridization of Active Contour Model and OFA. Journal of Intelligent Systems, 29(1), 35-52. https://doi.org/10.1515/jisys-2017-0415

[12] Schwarz Schuler, Joao Paulo & Romaní, Santiago & Abdel-nasser, Mohamed & Rashwan, Hatem & Puig, Domenec. (2021). Reliable Deep Learning Plant Leaf Disease Classification Based on Light-Chroma Separated Branches. 10.3233/FAIA210157.

[13] Vijai Singh, A.K. Misra,
Detection of plant leaf diseases using image segmentation and soft computing techniques, Information Processing in Agriculture, Volume 4, Issue 1, 2017, Pages 41-49, ISSN 2214-3173, https://doi.org/10.1016/j.inpa.2016.10.005.

[14] ImageJ documentation, https://imagej.net/plugins/mixture-modeling-thresholding, accessed on 18th May 2022.

[15] Nichole Gauthier (2018). "Apple scab" https://www.apsnet.org/edcenter/disandpath/fungalasco/pdlessons/Pages/AppleScab.aspx (2018)