Passion Fruit Disease Detection using Image Processing

Research · June 2019

DOI: 10.13140/RG.2.2.28353.20328

2 authors:

Deshan Dharmasiri
University of Kelaniya

Shantha Jayalal
University of Kelaniya

Passion Fruit Disease Detection using Image Processing

Dharmasiri, S.B.D.H.[1*], Jayalal, S.[1]
Department of Industrial Management
University of Kelaniya
Sri Lanka
*sbdeshan@gmail.com**

Abstract

*Fruit diseases are a major problem in economic losses and production in the agricultural industry worldwide. In this paper, an image processing approach is proposed for identifying passion fruit diseases. According to the Sri Lankan context, treatment details are taken by the farmers from the field officers. But it can take a few days. So, this proposed system can be used to identify passion fruit diseases quickly and automatically. This proposed approach is composed of the following main steps; Image Acquisition, Image Preprocessing, Image Segmentation, Feature Extraction, Dataset Preparation, Training & Testing. Healthy and two types of passion fruit diseases namely passion fruit scab and woodiness images, were used for this approach. This approach was tested according to passion fruit disease type and its' stages, such as mild, moderate and severe. K-Means clustering was used for segmentation. Images were clustered according to k values, such as 2, 4, 6 and 8. Before the segmentation, images were converted to RGB, L*a*b, HSV and Grey colour models, because of find out the most suitable colour model for this approach. Local Binary Pattern was used for feature extraction and Support Vector Machine was used for creating the model. 70% of each dataset was used to train the SVM and the other 30% was used to test the model. According to this approach, passion fruit diseases can be identified in the average accuracy of 79% and its' stage can be identified in average accuracy 66%.*

*Keywords: K-Means Clustering, Local Binary Pattern, Support Vector Machine, L*a*b*

1. Introduction

Passion fruit (*Passiflora edulis*) is a very famous fruit in modern society. Because of its nutrients and health benefits, such as prevention of cancer, controlling blood pressure and preventing hyperlipidemia. It has a huge demand in both local and international markets. Passion fruit is cultivated in countries like Australia, New Zealand, Kenya, South Africa, South America, Hawaii, India and Sri Lanka. Currently, South America has become the largest passion fruit producer in the world [1]. The annual passion fruit production is around 0.8 million MT in the world.

Nowadays many people tend to cultivate passion fruit in Sri Lanka, because of favourable climate conditions, temperature and attractive market demands. There are two variety of passion fruits in Sri Lanka. One is yellowish and another one is purplish. However, the yellowish passion fruit variety is widely cultivated in Sri Lanka, which is sourer than the purplish one [2]. Kalutara, Gampaha, Ratnapura and Kurunegala are the districts, which are popular in cultivating passion fruit. The average annual passion fruit production is 500MT in Sri Lanka. In 2017, total passion fruit cultivation extent was 470ha in Sri Lanka. Figure 1 shows the total passion fruit production in each previous years according to census and statistics data.

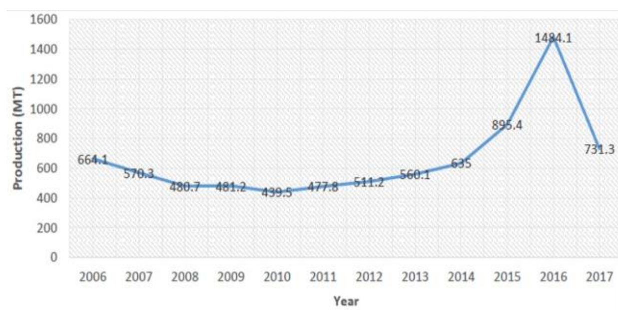

Figure 1. Passion fruit production in previous years

Figure 1 shows how the production of passion fruit in Sri Lanka since 2006. The production of passion fruit was steadily grown between 2006 and 2014. An average 500MT can be observed during this period. In 2014, there was a 635MT production and it increased sharply until 2016 when the production of passion fruit reached to 1484MT. It was the maximum production between 2006 and 2017. After 2016, production declines suddenly. In 2017, passion fruit production was 731MT from 470.27ha extent. When considering the ratio of production and hectare in 2016 and 2017, there was the 2.92 (MT per ha) in 2016. It was 1.56(MT per ha) in 2017. The decline of this ratio is 1.36(MT per ha). When considering only the production, it was around 50.7% decline in 2017.

This production loss was around Rs.90 million in 2017 [3]. Fruit Research and Development Institute (FRDI) mentions “Pests, diseases and bad climate mainly caused to decline the production in 2017”. After spreading diseases, farmers have to use fungicides, remove infected parts and burn them with a systematic way. So initial disease identification is very important to prevent spreading diseases. However, this prevention cannot be successful due to the lack of agricultural field officers.

Nowadays technology plays a vital role in all the fields but till today traditional methodologies are used for agriculture in Sri Lanka. However the large scale of agricultural countries use technologies like MRI, X-ray imaging for detecting the quality of the fruits. But these technologies are costly for farmers to afford, occupy large space, users need to have the knowledge to use and analyze the results.

However, today passion fruit disease identification is done manually by 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 passion fruit diseases accurately.

2. Literature review

Recently, many people have done researches for detecting fruit and vegetable diseases using image processing and deep learning. According to the research paper [4], authors had used image processing technology to identify the pomegranate diseases. Image preprocessing was the first step of the methodology. Image resizing was done under the image preprocessing. Because digital camera had been used to capture the images in this study. The size of those images was very large and take more time to process. So all images were resized to 300 x 300 PX. Morphology, colour and CCV features were used for feature extraction. K-means clustering technique was used for partitioning the training dataset according to their features. After the clustering, SVM was used for classification to identify the image as infected or non-infected. An intent search technique was provided to find the user intention. The best result was got using morphology feature extraction. Experimental evaluation of this approach was effective and 82% accurate to identify pomegranate disease.

In paper [5], the authors presented the image processing based approach for fruit disease detection. First, read input image and transformed it from RGB to L*a*b colour space. Because the colour information in the L*a*b colour space is stored in only two channels. Input images were partitioned into four segments using K-means cluster in this research. Because the empirical observations it was found that using 3 or 4 clusters yield good segmentation results. GCH, LBP, CCV and CLBP were used for feature extraction. More accurate results could be taken using CLBP feature extraction technique. K-means clustering was used for segmentation. Those segmented images were extracted to label each pixel in the image. SVM algorithm was used for training and classification of fruit disease. Authors used apple as a test case and evaluated the classification model for three types of apple diseases, which were apple rot, apple blotch and apple scab. The accuracy of this approach was achieved by up to 93%.

In paper [6], the author had used SVM classification for identifying and classifying the grape leaf diseases. Grape leaf images were taken using a digital camera and those were used to both training and testing the system. Collected images included the leaves infected by Powdery Mildew and Downy Mildew. Removing background noise and resizing to 300*300 PX to improve the image quality were done under the image preprocessing. Gaussian filtering had been used to remove noise in the image. K-means clustering was used for segmenting an image into three groups. Features were extracted based on both colour and texture for taking accurate disease information. Finally, the classification model was used to detect the leaf disease. LSVM was used in this research for the classification of leaf diseases. This system could detect and classify the examined disease successfully. The accuracy of this system was 88.89%.

In paper [7], authors had used image processing technology for identifying the leaf diseases. First authors selected the plants, which were affected by the disease and then took the snapshot of the diseased leaf. Contrast enhancement and converting RGB to HIS was done under the image preprocessing step. K-means clustering algorithm was used to cluster the object based on the feature of leaf into k number of groups. SVM algorithm had been used in this system for classification purpose. SVM is a statistical learning-based solver. Finally, when entered a diseased leaf image to a system, the system was able to detect the leaf disease successfully.

In paper [8], authors had presented the image processing based system to identify pomegranate fruit diseases. This fruit is mainly affected by Bacterial Blight, Anthracnose and Alterneria. After capturing the disease images, image resizing, filtering, segmentation, morphological features were used to preprocess the images. Image segmentation is the process of dividing the image into multiple parts. Colour-based segmentation was used in this research, such as clustering, YCbCr, RGB, L*a*b and HSV. However, the best performance in terms of segmentation error was achieved by the HSV and YCbCr. Morphology, texture and colour features were extracted for classification purpose. HIS colour model and colour histogram techniques had been used to colour feature extraction. Under the morphology feature extraction, boundary extraction was used to identify the region and shape. The eroded images were subtracted from the original image to extract a shape from healthy fruit image. Gabor filter was used to texture feature extraction. After the training and testing of images, diseased and non-diseased fruits were classified using Minimum Distance Classifier (MDC).

3. Methodology

There are six phases in this methodology. Those are Image Acquisition, Image Preprocessing, Image Segmentation, Feature Extraction, Dataset Preparation, Training and Testing.

3.1. Image Acquisition

In this phase, the sample images are collected, which are required to train the classifier algorithm and build the classifier model. Yellowish passion fruit variety was selected to take sample images. Because the yellowish variety is widely cultivated in Sri Lanka. Healthy and diseased passion fruit images were taken by using 13MP mobile phone digital camera and used for both training and testing the classifier algorithm. Images were taken in different angles, under the different environmental and lighting conditions. The standard JPG format was used to store these images. In this study, images were collected from farms in different regions like Monaragala, Gampaha and Kaluthara. Because according to the statistic and census data, high passion fruit production showed in these districts in 2015. Passion fruits infected by scab disease and woodiness virus, that had been included in collected images. That's why scab disease and woodiness virus were selected for training and testing processes. Images are shown by figure 2, figure 3, and figure 4, which are the categories of passion fruits that are collected under the data acquisition.

Figure 2. Healthy passion

Figure 3. Scab disease

Figure 4. Woodiness virus

3.2. Image Preprocessing

After the image acquisition, image processing was done for improving the image quality. All original passion fruit images were stored in one folder, called "Images_Original". Those images were named as img_x. "x" can take any value of numbers. Only horizontal images were rotated by 90 degrees and resized by 200x300 pixels. Vertical images were resized by 200x300 pixels and when the width and height of the image are same, those images were resized to 250x250 pixels. When the image size is too large, the processing task takes more time [4]. After that, one of the noise reduction methods was used to remove the noises from images and increase the sharpness of images. Later, all preprocessed images were saved in a folder, called "Images_Preprocessed".

3.3. Image Segmentation

The third phase of the methodology is image segmentation. As the first step, all preprocessed images were converted into L*a*b, HSV, Grey colour models and kept one in the original way (RGB). Because the identifying suitable colour model for preprocessing is one of the outcomes of this research. After converting them by using colour models, those images were renamed as mention in table 1.

Table 1. Image naming according to colour model

Colour model	After using colour models
RGB	img_x_0
L*a*b	img_x_1
HSV	img_x_2
Grey	img_x_3

After that, the image was converted to float32 format matrix. This matrix values were clustered using the K-means clustering algorithm. Four values were taken as the k value for this clustering mechanism. Such as 2, 4, 6 and 8. Clustered matrix values were converted to an image. According to the k values, clustered images were renamed as the table 2 and each image was saved in a specific folder. An L*a*b coloured image is mentioned in this following table 2 as the example.

Table 2. Image naming according to clustering

K value	After clustering
2	img_x_1_2
4	img_x_1_4
6	img_x_1_6
8	img_x_1_8

Figure 5 shows the sample images, which were received after the image segmentation.

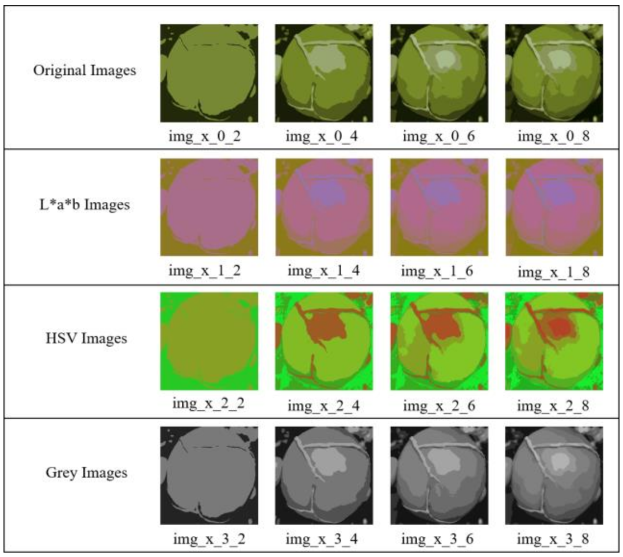

Figure 5. Image segmentation result images

3.4. Feature Extraction

The fourth phase of the methodology is feature extraction. Local Binary Pattern (LBP) mechanism was used for feature extraction. Segmented images were taken from folders one by one and each image was converted to grey-scale. Then a neighbourhood was selected for each pixel in grey-scale image surrounding the center. An LBP value was calculated for this center pixel and stored the output 2D array with the same width and height as the taken image.

Then calculating the LBP for center pixel was started from top right neighbourhood pixel clockwise. The results of binary values were stored in an 8-bit array, which was converted to decimal later. These converted decimal values were stored in the output LBP 2D array. A histogram for each image was computed using the output 2D array. For this implementation, the 3x3 pixel combination was used. So 256 (2^8) number of feature histogram values were received at the end. Those histogram values were used for creating the datasets. Feature extracted values of RGB, L*a*b, HSV and Grey images were taken into comma-separated value (CSV) file. So at the end, four CSV files were created.

3.5. Dataset Preparation

The fifth phase of the methodology is datasets preparation. Labels could be added to CSV files, which were created using feature extraction. However, three image sets were created to do experiments. Preparation of those image sets is discussed here. Field expertise support was taken for the categorization of images and each image were selected from the categorized sets of an image randomly.

3.5.1. Considering disease name only

There are three categories in this image sets. Those are non-disease, scab disease and woodiness virus. A number of images were taken into each category is mentioned in the below tables. Table 3 shows the labels, which were assigned for each category. Table 4 shows the count of images, which were included in each dataset.

Table 3. Labelling according to disease name

Non-disease	Scab disease	Woodiness virus	Mild	Moderate	Severe	Label
1	0	0	0	0	0	100000
0	1	0	0	0	0	010000
0	0	1	0	0	0	001000

Table 4. Three classes of datasets image amounts

Disease name	Non-disease	Scab disease	Woodiness Virus
No. of images (1st dataset)	13	34	40
No. of images (2nd dataset)	26	67	80
No. of images (3rd dataset)	39	98	120

3.5.2. Considering disease names and its stages

Stages of the diseases are considered here. There are three types of stages, such as mild (I), moderate (II) and severe (III). According to those stages, there are seven categories in this image sets. Those are non-disease, mild-scab, moderate-scab, severe-scab, mild-woodiness, moderate-woodiness and severe-woodiness. Table 5 shows the labels, which were assigned for each category. Table 6 shows the count of images, which were included in each dataset.

Table 5. Labelling according to diseases' stage

Non-disease	Scab disease	Woodiness virus	Mild	Moderate	Severe	Label
1	0	0	0	0	0	100000
0	1	0	1	0	0	010100
0	1	0	0	1	0	010010
0	1	0	0	0	1	010001
0	0	1	1	0	0	001100
0	0	1	0	1	0	001010
0	0	1	0	0	1	001001

Table 6. Seven classes of datasets image counts

Disease	Non-disease	Scab Disease			Woodiness Virus		
Stage		I	II	III	I	II	III
1st set	13	12	12	10	13	15	12
2nd set	26	24	23	20	26	30	24
3rd set	39	36	33	29	39	45	36

These image sets were preprocessed, segmented and feature extracted separately. Four CSV files were received for every image set. Original image name and label included into a one CSV file, which called a "label file". This file was used for giving labels for each row in feature extracted CSV file. Name column in feature extracted and label files were matched. When similar names were found, a particular label in label file was given to label column in the feature extracted file. After that, this dataset can be used for training and testing the model.

A sample feature CSV file is shown in figure 6. Image name, feature histogram values and labels are included in this dataset.

Image Name	V_1	V_2	....	V_255	V_256	Label
Img_1_3_2	882	155	....	421	51975	100000
Img_10_0_4	1629	278	....	750	42711	010000
Img_2_1_2	1938	274	....	886	37969	001000
Img_50_2_6	2128	291	....	989	33197	010100
Img_99_3_8	988	175	....	482	50535	010010
Img_14_2_2	1698	262	....	702	32531	001100
Img_55_3_4	995	167	....	900	33173	001010

Features histogram values

Figure 6. Sample dataset

3.6. Training & Testing

The last phase of the methodology is training and testing the model. Support Vector Machine (SVM) algorithm was used for classification as mentioned in the methodology. A set of mathematical functions are used in SVM algorithms, which are defined as kernels. The function of the kernel is to take data as input and transform those data into the required form. SVM algorithms are used with different types of kernels. Gaussian, sigmoid, hyperbolic tangent and polynomial kernel are some of them. Among those kernels, the polynomial kernel is popular in image processing. So, the polynomial kernel was used for the SVM algorithm as the kernel in this study. Each CSV files were used for training the SVM separately. 70% of the data from each dataset was used for training purpose. Every training time, CSV file rows were shuffled for increasing the accuracy. When considering the testing the model, the remaining 30% of data from each dataset were used. In here, the model predicted the label for each image and those were evaluated using actual labels. The accuracy of these models can be defined as,

$$Accuracy\ (\%) = \frac{\text{Total number of images correctly classified}}{\text{Total number of images used for testing}} * 100$$

4. Experimental Results

While labelling the features CSV file, three labels were used for identifying passion fruit disease according to its name. These CSV files are called as "3 classes dataset". Another way is using seven labels for features CSV file to identify passion fruit diseases according to its stage. This CSV files are called as "7 classes dataset". Every training and testing time, rows of CSV files were shuffled randomly for increasing the accuracy of the model. Each CSV file was trained and tested in five times and accuracy was taken. Average of those accuracies was taken as the accuracy of each model. Figure 7 shows the total number of images, which were used to create each dataset.

Using this image dataset, four types of CSV files were created. Such as Grey, HSV, L*a*b and RGB, which are the colour models that were used to evaluate in this study.

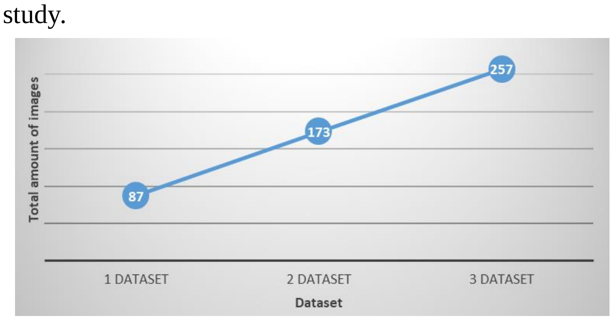

Figure 7. Total number of images in each dataset

4.1. Results According to 3 Classes Datasets

4.1.1. First dataset results

348 image records are included in these CSV files and labelled by 3 labels. Each CSV file was used five times for getting accuracy and by using those accuracies, got the average accuracy of the model. Those models accuracies are discussed in table 7.

Table 7. Accuracy of models using 3 classes of first dataset

	1st time	2nd time	3rd time	4th time	5th time	Average Accuracy
Grey	76.9	75.0	68.3	71.2	69.2	72%
HSV	67.3	75.9	69.2	72.1	69.2	70%
L*a*b	78.9	73.1	78.9	81.7	81.7	79%
RGB	70.2	69.2	75.0	63.5	71.2	70%

4.1.2. Second dataset results

692 image records are included in these CSV files and labelled by 3 labels. Each CSV file was used five times for getting accuracy and by using those accuracies, got the average accuracy of the model. Those models accuracies are discussed in table 8.

Table 8. Accuracy of models using 3 classes of second dataset

	1st time	2nd time	3rd time	4th time	5th time	Average Accuracy
Grey	63.9	68.3	69.2	74.0	67.3	68%
HSV	65.4	59.1	52.4	65.4	60.9	60%
L*a*b	72.1	73.6	75.5	69.2	69.7	72%
RGB	71.2	67.3	67.3	67.3	73.6	69%

4.1.3. Third dataset results

1028 image records are included in these CSV files and labelled by 3 labels. Each CSV file was used five times for getting accuracy and by using those accuracies, got the average accuracy of the model. Those models accuracies are discussed in table 9.

Table 9. Accuracy of models using 3 classes of third dataset

	1st time	2nd time	3rd time	4th time	5th time	Average Accuracy
Grey	58.8	59.4	65.9	61.7	60.7	61%
HSV	56.1	58.4	56.5	54.9	57.4	56%
L*a*b	68.5	64.3	61.4	65.9	67.2	65%
RGB	58.8	64.3	60.4	58.1	59.4	60%

4.2. Results According to 7 Classes Datasets

4.2.1. First dataset results

348 image records are included in these CSV files and labelled by 7 labels. Each CSV file was used five times for getting accuracy and by using those accuracies, got the average accuracy of the model. Those models accuracies are discussed in table 10.

Table 10. Accuracy of models using 7 classes of first dataset

	1st time	2nd time	3rd time	4th time	5th time	Average Accuracy
Grey	64.4	63.5	57.7	64.4	59.6	62%
HSV	63.5	57.7	55.8	52.9	56.7	57%
L*a*b	65.4	66.4	68.3	70.2	62.5	66%
RGB	63.5	63.5	59.6	70.2	61.5	64%

4.2.2. Second dataset results

692 image records are included in these CSV files and labelled by 7 labels. Each CSV file was used five times for getting accuracy and by using those accuracies, got the average accuracy of the model. Those models accuracies are discussed in table 11.

Table 11. Accuracy of models using 7 classes of second dataset

	1st time	2nd time	3rd time	4th time	5th time	Average Accuracy
Grey	53.4	54.8	49.5	51.4	56.7	53%
HSV	45.7	48.6	51.9	49.0	52.9	50%
L*a*b	60.6	61.1	62.0	58.7	60.6	61%
RGB	61.5	59.1	58.2	58.7	52.9	58%

4.2.3. Third dataset results

1028 image records are included in these CSV files and labelled by 7 labels. Each CSV file was used five times for getting accuracy and by using those accuracies, got the average accuracy of the model. Those models accuracies are discussed in table 12.

Table 12. Accuracy of models using 7 classes of third dataset

	1st time	2nd time	3rd time	4th time	5th time	Average Accuracy
Grey	47.1	49.7	53.3	53.9	49.7	51%
HSV	47.4	40.3	41.2	43.8	47.1	44%

L*a*b	56.8	54.9	63.6	58.8	62.3	59%
RGB	47.4	44.2	45.8	51.9	47.4	47%

4.3. Accuracy Comparison

4.3.1. According to 3 classes datasets

Figure 8 shows the accuracy comparison according to 3 classes datasets. L*a*b gives the highest accuracies and HSV gives the lowest accuracies among the preprocessing colour models. Grey and RGB take a similar place in figure 8 when considering as overall. According to figure 8, the accuracy of each model decrease within the number of images in each dataset.

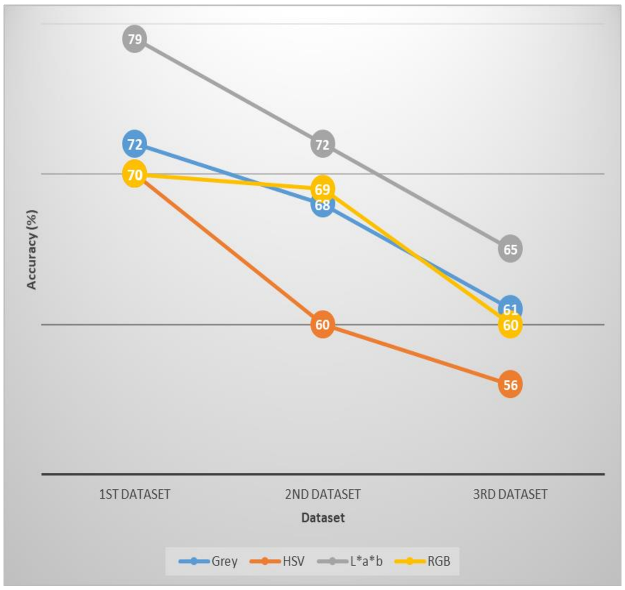

Figure 8. Accuracy of models using 3 classes datasets

4.3.2. According to 7 classes datasets

Figure 9 shows the accuracy comparison according to 7 classes datasets. L*a*b gives the highest accuracies and HSV gives the lowest accuracies among the preprocessing colour models. When considering the grey colour model, it takes the second place of the accuracy comparison chart. The third place of accuracy comparison chart is taken by the dataset, which images were not used any preprocessed colour model. That is the RGB colour model. According to figure 9, the accuracy of each model decreases, when increasing the number of images in each dataset.

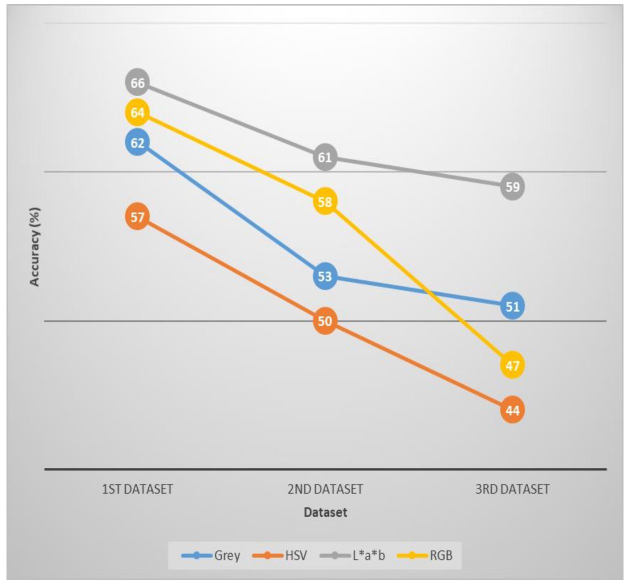

Figure 9. Accuracy of models using 7 classes datasets

5. Conclusions

In this study, the support vector machine algorithm was used for creating the models, which were built according to passion fruit diseases and its stages. The passion fruit diseases can be identified in the average accuracy of 79% and its stage can be identified in the average accuracy of 66%.

When considered the 3 classes datasets, the highest average accuracy of the model was received by the first dataset, which included non-disease (13), scab disease (34) and woodiness virus (40). A number of images in each category has been mentioned in parenthesis. There are 87 images were included in this dataset. The accuracy was 79%, which was received using L*a*b colour model.

When considered the 7 classes datasets, the highest average accuracy of the model was received by the first dataset, which also included 87 images. Those images were categorized as non-disease (13), mild-scab (12), moderate-scab (12), severe-scab (10), mild-woodiness (13), moderate-woodiness (15) and severe-woodiness (12). A number of images in each category has been mentioned in parenthesis. The model had 66% accuracy and it was received by the L*a*b colour model.

6. References

[1] Sema, A. & Maiti, S. C. *Status & Prospects of Passion Fruit Industry in Northeast India.* Information bulletin, Central Institute of Horticulture, Medziphema, Nagaland, 2009

[2] Kalubowila, I., David, S. M. G. D., & Rathnayaka, R. M. M. N. K. *Improvement of Yellow Passion Fruit Through Recurrent Selection.* Sri Lanka, Annals of Sri Lanka Department of Agriculture, 2017

[3] Central Bank of Sri Lanka. *Annual Report 2017,* Colombo 01: Central Bank of Sri Lanka,

[4] Bhange, M. & Hingoliwala, H. A. *Pomegranate Disease Detection Using Image Processing.* India, Elsevier B.V, 2015

[5] Dubey, S. R. & Jalal. A. S. *Adapted Approach for Fruit Disease Identification using Images.* India, International Journal of Computer Vision and Image Processing, 2012

[6] Padol, P. B. *SVM Classifier Based Grape Leaf Disease Detection.* India, Conference on Advances in Signal Processing (CASP), 2016

[7] Sujatha, R., Kumar, Y. S., Akhil, G, U. *Leaf disease detection using image processing.* Vellore, Journal of Chemical and Pharmaceutical Sciences, 2017

[8] Khot, S. T., Supriya, P., Gitanjali, M., & Vidya, L. *Pomegranate Disease Detection Using Image Processing Techniques.* Pune, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2016