

PHYTOPLASMA DISEASE OF CASSAVA LEAF

CLASSIFICATION AND ANALYSIS

USING COLOR FEATURES

&

LINEAR SUPPORT VECTOR MACHINE

Niño M. Austria

Csci 142 - Research Presentation

1 . I N T R O D U C T I O N

1.1 Background of the Study

Cassava, scientifically known as "Manihot esculenta," holds significant importance worldwide due to its versatile nature and nutritional value. According to tuasaude.com, cassava helps to prevent cardiovascular diseases, constipation and some types of cancer. A good source of vitamin which promotes skin and hair health, aid weight loss and many more.



Nutritional Information for 100 grams of Cassava

Components	100 g (3 tablespoons) of boiled cassava
Energy	125 calories
Protein	0.6 g
Carbohydrates	30.1 g
Fat	0.3 g
Fiber	1.9 g
Vitamin C	18.2 mg
Vitamin A	13 mcg
Carotenoids	13 mcg
Folic acid	24 mcg
Calcium	19 mg
Potassium	100 mg
Magnesium	27 mg



1.1 Background of the Study

While cassava roots are widely recognized, it's the often overlooked, nutrient-rich cassava leaves that truly deserve the spotlight.

According to a review on the potential of cassava leaves in human nutrition, Cassava leaves are rich in nutrients favourably comparable with other green vegetables which may be used as an important supplement to prevailed starchy diets, if properly detoxified. Cassava leaves are used as food in various countries of the world.

However, there are diseases that kill the cassava plant, most likely attacking the leaves. These diseases include Mosaic Disease, Brown Streak Disease, Bacterial Blight, Phytoplasma, and many more.



As a concerned citizen, what will you do to help against this problem?

1.2 Objective of the Study

Generally, this study aims to utilize computer vision to develop an efficient model for classifying the leaves of cassava plants to determine whether they are healthy or infected by phytoplasma.

Specifically, it aims to,

- implement and analyze the effectiveness of color features in classifying phytoplasma disease in cassava leaves.
- examine the utilization of linear support vector machine in categorizing phytoplasma-infected cassava leaves.



1.3 Significant of the Study

By developing a classification model and having detailed analysis to classify cassava leaves as healthy or infected:

- optimizes resource allocation within farming communities, minimizing wastage and maximizing productivity
- offer cost-effective disease monitoring and management, promising significant improvements in crop productivity and livelihoods.
- represents a convergence of technology and agriculture, poised to contribute to sustainable farming and global food security.



1.4 Scope and Limitations

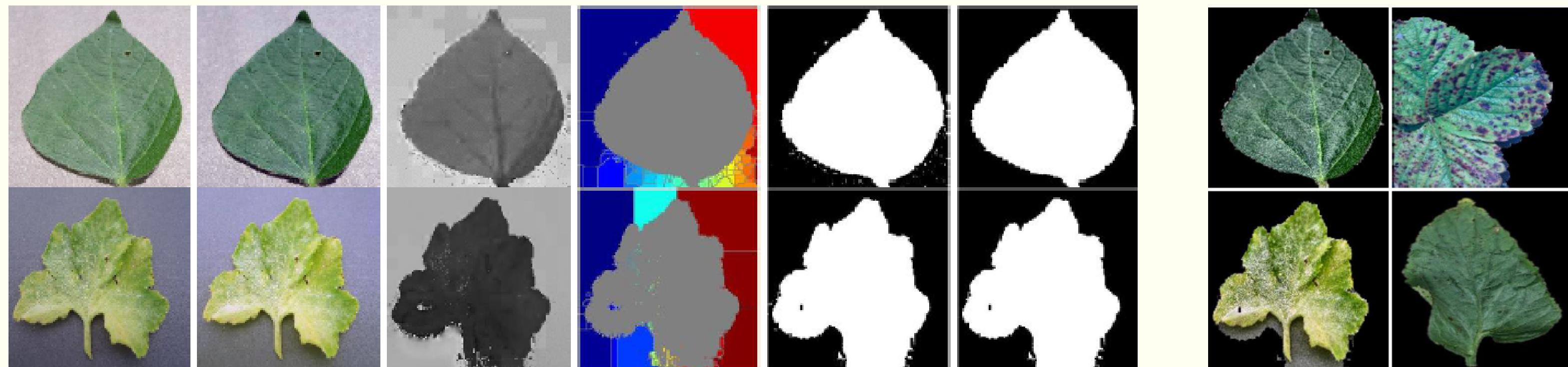
This study will only cover classification and analysis between healthy leaves of cassava plants and those infected by phytoplasma disease. The datasets were collected from various areas where cassava plants are present in Baybay City, Leyte. Then, it is limited to the development of a classification model for healthy and infected leaves. It includes experimentation with image processing techniques, extraction and analysis of different color-based features, and the implementation of intelligent system.



2 . R E L A T E D W O R K

Based on the study entitled "Leaf Image based Plant Disease Identification using Color and Texture Features", identification of plant disease is usually done through visual inspection or during laboratory examination which causes delays resulting in yield loss by the time identification is complete. The proposed approach contributes automated identification of plant diseases which follows a sequence of steps involving pre-processing, segmentation of diseased leaf area, calculation of features based on the Gray-Level Co-occurrence Matrix (GLCM), feature selection and classification. In the said study, six color features and twenty-two texture features have been calculated. The proposed model of disease identification provides an accuracy of 98.79% with a standard deviation of 0.57 on 10-fold cross validation. The accuracy on a self-collected dataset is 82.47% for disease identification and 91.40% for healthy and diseased classification.

PREPROCESSING STEPS AND BACKGROUND REMOVAL



Unprocessed

Color Adjusted

Hue Layer

Watershed -
Transformed

Segmented Mask

Morphological
Processed

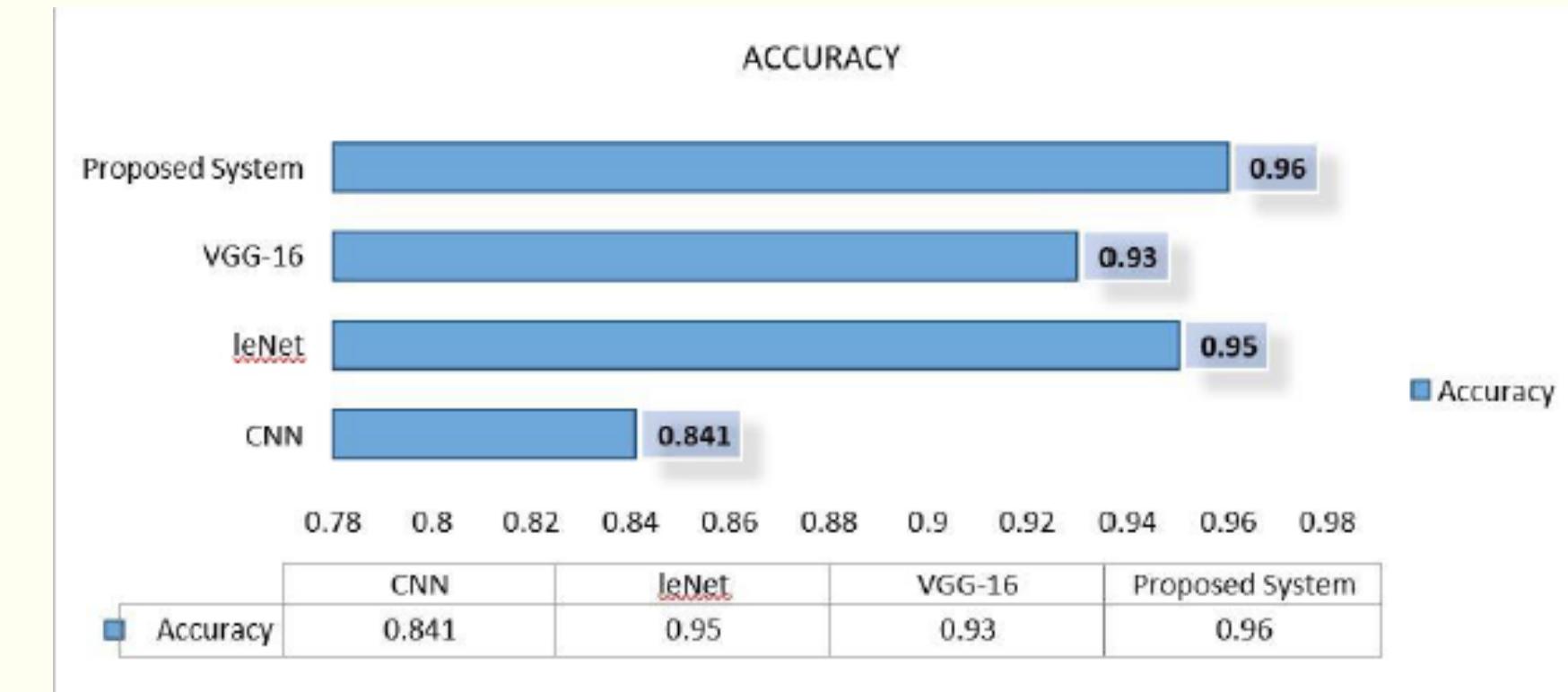
After Background Removal



Based on the study entitled "Optimal Plant Leaf Disease Detection using SVM classifier with Fuzzy System," the proposed SVM classifier exhibited superior performance in terms of accuracy compared to other intelligent systems.



Plant Leaves Datasets

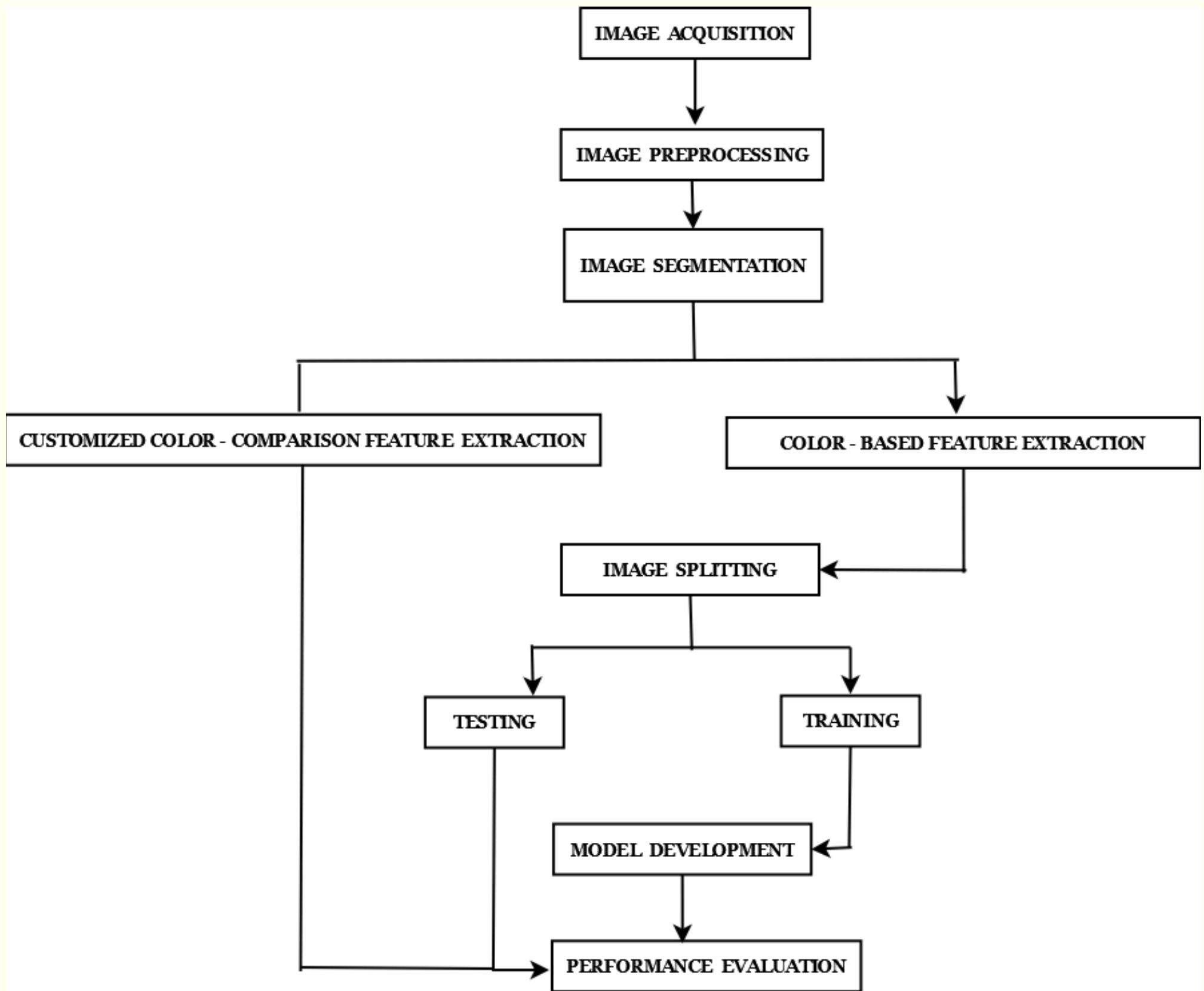


Accuracy Comparison between the Proposed System and Other Intelligent Systems



3 . M E T H O D O L O G Y

Proposed Methodology



3.1 Image Acquisition

1. An area where Cassava plants were present in Baybay City, Leyte, was identified.
2. Collaboration was established with an expert Instructor Michelle E. Gumba from the VSU Department of Plant Breeding and Genetics
3. Permission was obtained from the plant owner to collect leaves (datasets)..
4. Each leaf was placed on a white background bond paper, and with the assistance of an artificial lamp, the leaf was captured using a mobile phone.
5. The captured images were saved as raw dataset files.



Camera: 108 + 13 + 2 MP rear / 50 MP front
Operating System: Android 13, XOS 13



I. Straight Position

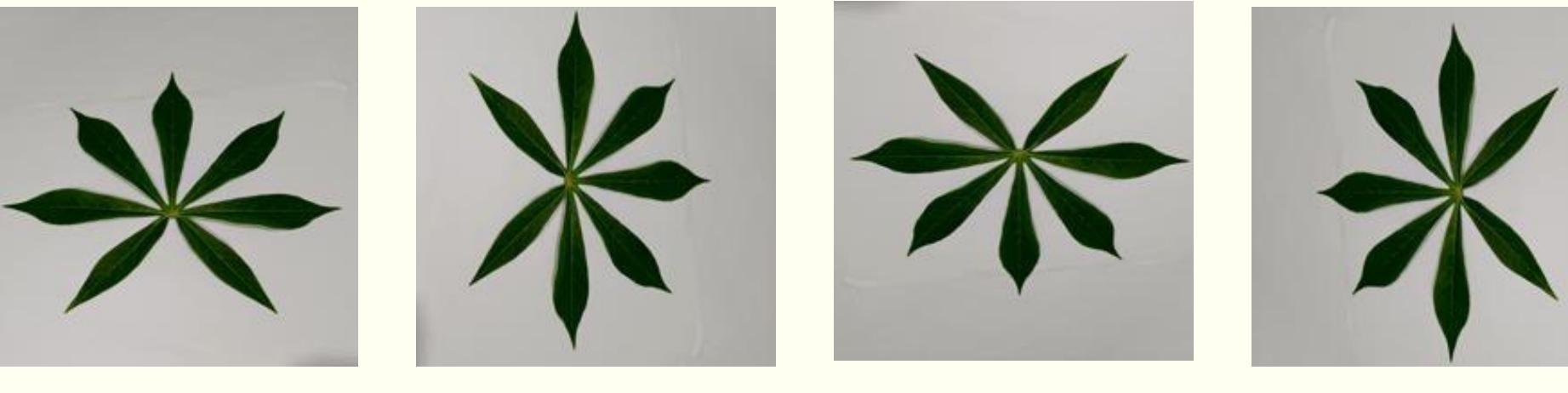


II. Slant Position



3.2 Image Preprocessing

1. The raw leaf dataset was rotated to generate images at 3 multiple angles.
2. Image resizing was applied to achieve dimensions of 200 x 200.



I. Healthy Datasets with rotated versions

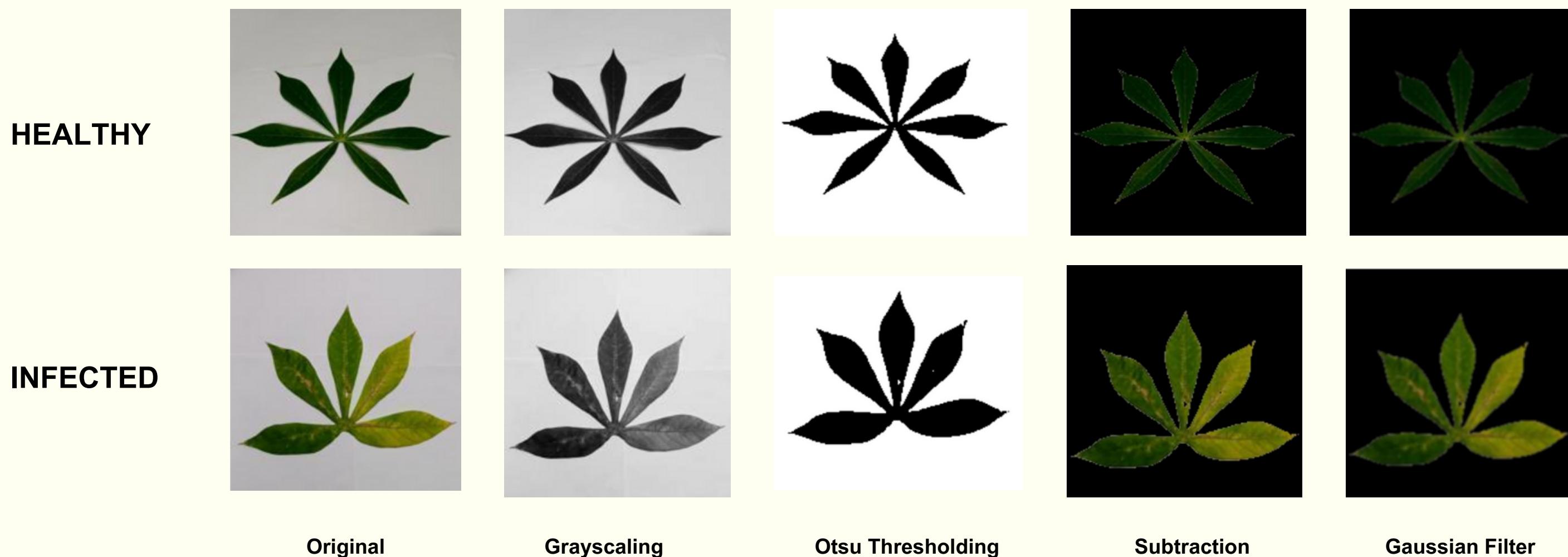


II. Infected Datasets with rotated versions

3.3 Image Segmentation

In the Background Removal Approach:

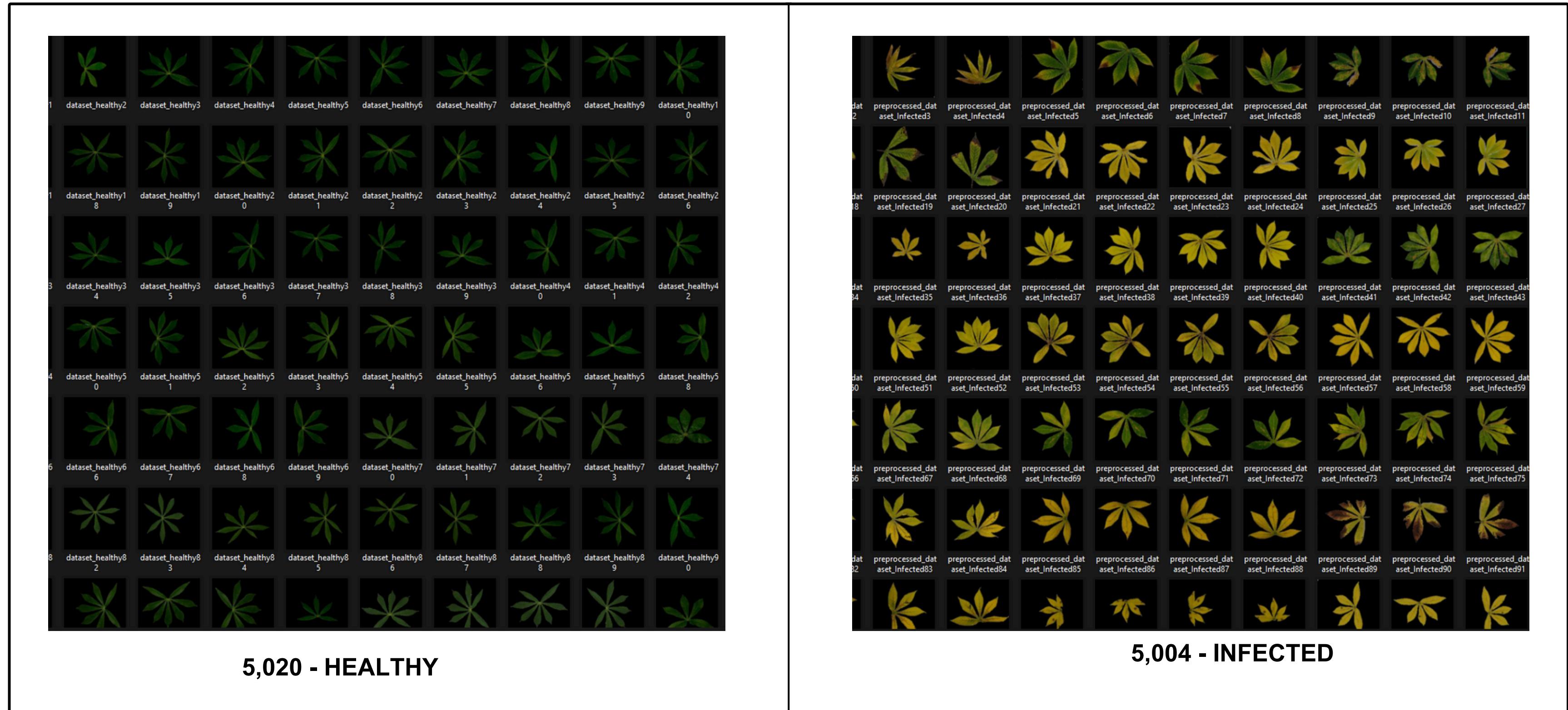
- Grayscale was applied.
- Otsu Thresholding was applied.
- Image Subtraction was applied.
- Gaussian Filter was applied to remove noise.



Step - By Step Process of Background Removal

3.3 Image Segmentation

DATASETS

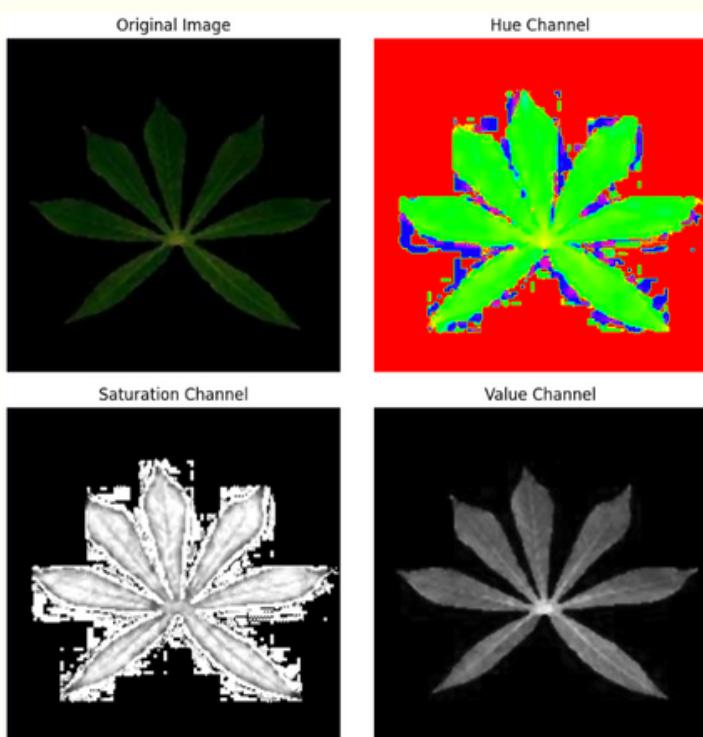


3.4 Feature Extraction

A. Customized Color - Comparasion Feature Extraction

If the cassava leaves turn yellow, covering more than 50% of the leaf and appearing yellower than healthy leaves, it indicates infection by phytoplasma disease.

In this approached, the image was converted to the HSV color space. The Hue channel was extracted, and thresholds for yellow and green colors in the Hue space were defined. The Hue channel was then thresholded to obtain yellow and green regions. Contours were found in the masks, and container arrays for yellow and green containers were initialized. Iteration was done through yellow contours to retrieve pixel coordinates and green contours to obtain pixel coordinates. It was checked whether the number of yellow pixels was greater than or lesser than the number of green pixels, and labeled appropriately. Experimentation and exploration beyond just yellow and green were encouraged, as those colors were primary indicators of healthy and infected cassava leaves. In this study, the comparison between the Hue channel and BGR channel approach was made using this feature extraction.



Piecewise Function

$$f(d) = \begin{cases} d = 1, & (y > g) \\ d = 0, & \text{otherwise} \end{cases}$$

Where:

d = dataset

g = [total computed green pixels of d]

y = [total computed yellow pixels of d]

0 = Healthy 1 = Infected

3.4 Feature Extraction

B. Global Color Histogram

GCH is the most known color histogram used to detect similar images. Feature extraction algorithm: Discretize your color-space (images' colors) into n color (You may use just $8*8*8=512$ color instead of $256*256*256=16777216$ color). The segmented datasets were extracted using Global Color Histogram.

In this study, the HSV color space was utilized. The image was converted to the HSV color space, and histograms were calculated for each channel (Hue, Saturation, Value). The histograms were concatenated and normalized. Features were extracted for healthy and infected images. A DataFrame was created to store the features and labels, and it was saved to an Excel file.

3.4 Feature Extraction

C. Color Moments

Color moments are measures that can be used to differentiate images based on their features of color. Once calculated, these moments provide a measurement for color similarity between images. These values of similarity can then be compared to the values of images indexed in a database for tasks like image retrieval.

1. Mean

- This represents the average color value of the dataset. For example, the mean would capture the average shade of green for healthy leaves and the average shade of yellow for infected leaves.

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i$$

2. Standard Deviation

- This measures the dispersion or spread of color values around the mean. A higher standard deviation indicates a wider range of colors present in the dataset.

$$\sigma \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

3.4 Feature Extraction

C. Color Moments

3. Skewness

- measures the asymmetry of the color distribution. A positive skewness value indicates that the distribution is skewed towards higher color values, while a negative skewness indicates skewness towards lower color values.

$$\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3$$

4. Kurtosis

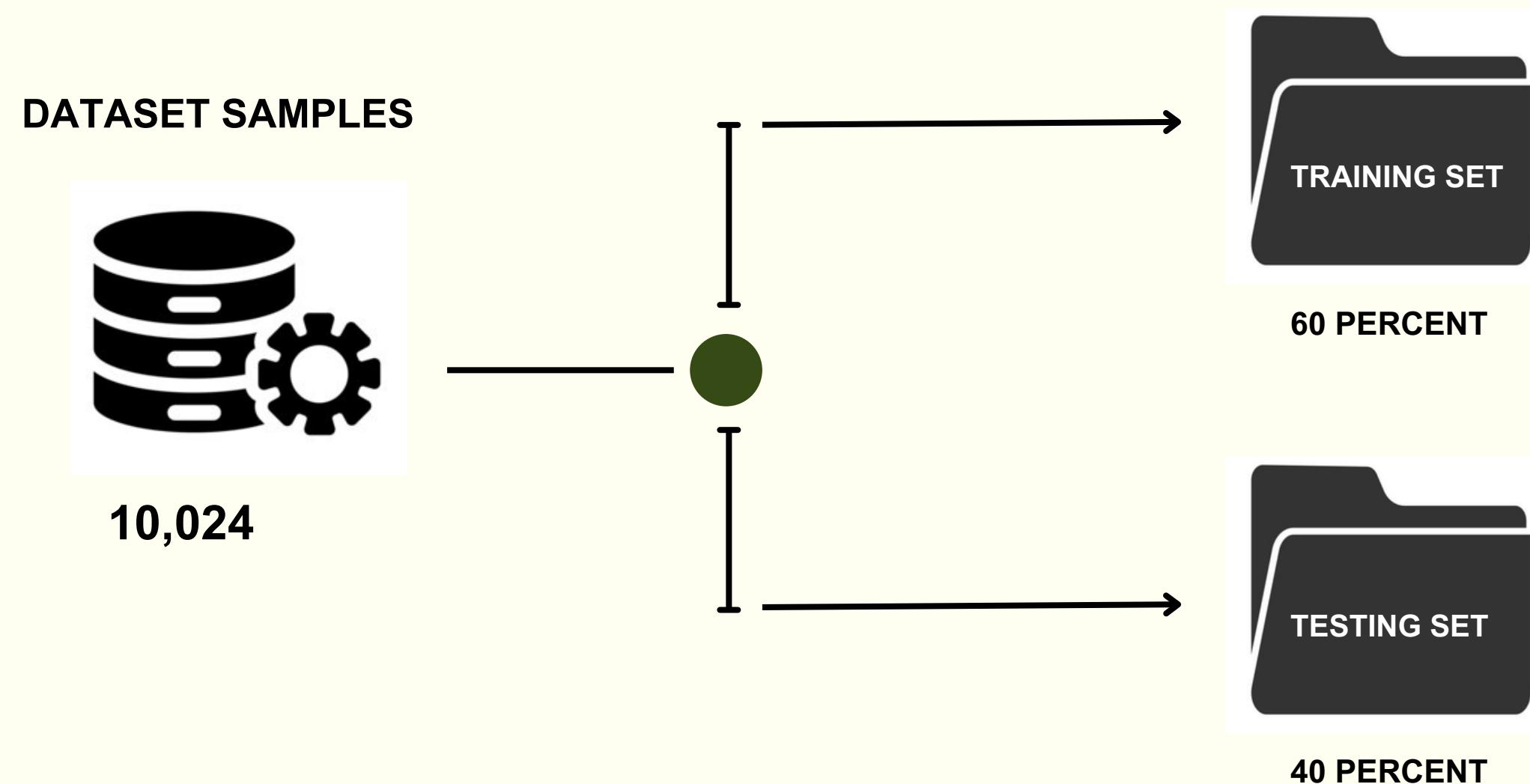
- measures the peakedness or flatness of the color distribution. A higher kurtosis value indicates a sharper peak (more concentrated distribution) compared to a normal distribution, while a lower kurtosis value indicates a flatter peak.

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$$

3.5 Image Splitting

The datasets were divided into training and testing sets for model development and evaluation.

Considering Cross Validation Techniques

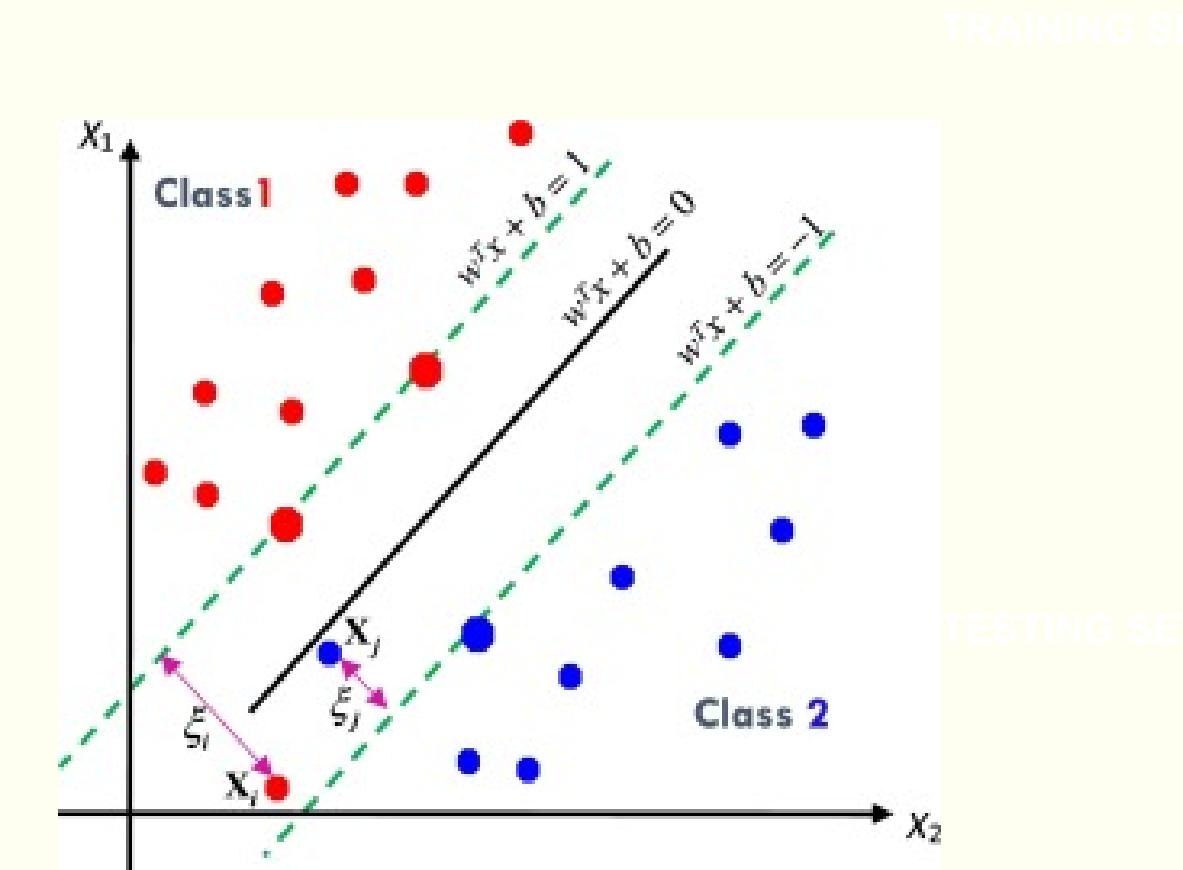


3.6 Model Development

After the datasets were prepared, the **[training datasets]** were trained and implemented to Linear Support Vector Machine.

Why Linear Support Vector Machine?

SVM works well when the classes are linearly separable, meaning they can be separated by a straight line in the feature space. Since the study is using color features for classification, it's likely that these features can be effectively separated by a linear boundary. It can handle a large number of features without overfitting, which is important when dealing with complex datasets like those in image processing.



3.7 Performance Evaluation

Forty percent (40%) of the datasets will be allocated for the testing phase. Out of these, four thousand and ten entries were utilized as testing data to evaluate the system's performance. Following this, cross-validation will be conducted to further assess the system's robustness. The results obtained will be visualized in a confusion matrix and graph. From this matrix, we can derive the classifier's accuracy and precision.

$$\text{Accuracy} = \frac{\text{sum of correctly classification}}{\text{total number of classifications}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

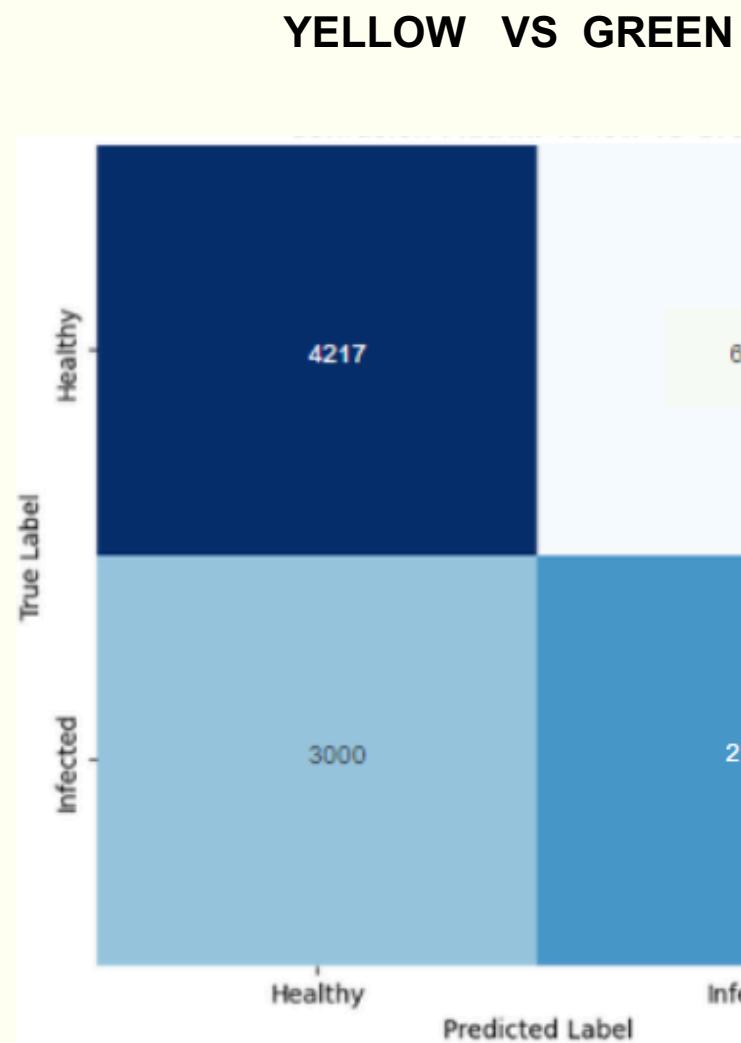


4 . R E S U L T S A N D D I S C U S S I O N

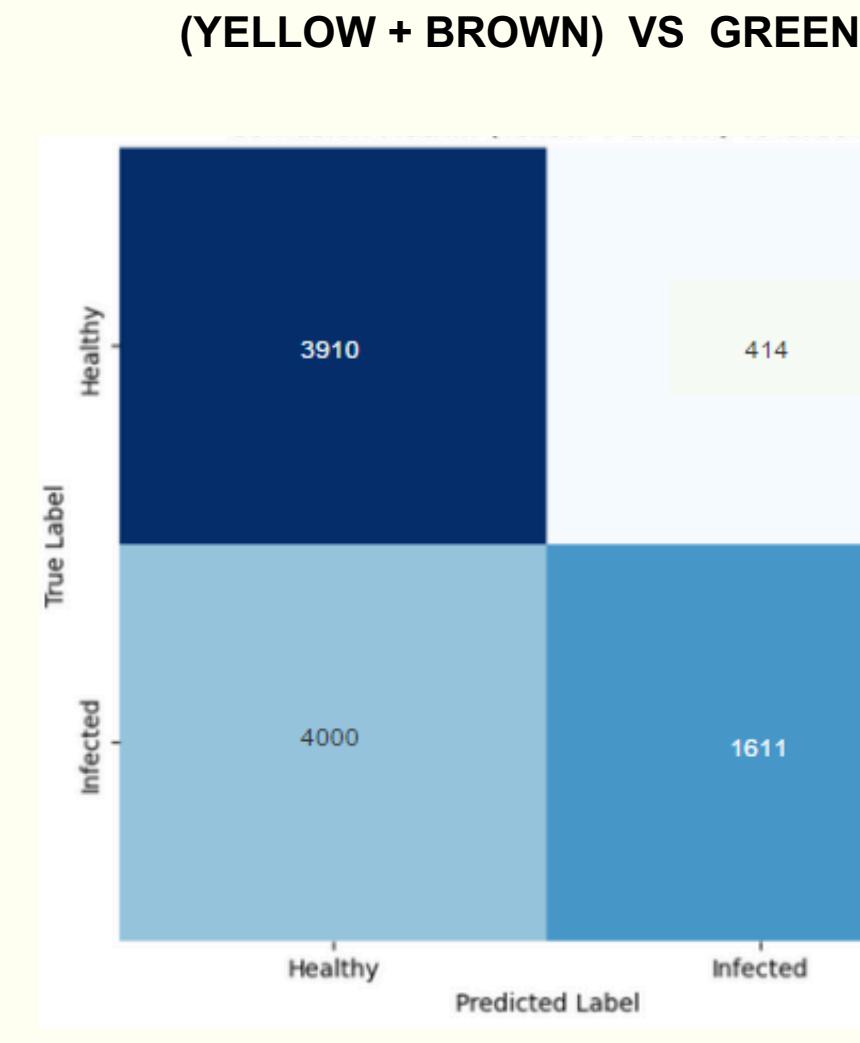
A. Customized Color - Comparasion Feature Extraction

USING RGB SPACE CONFUSION MATRIX RESULTS

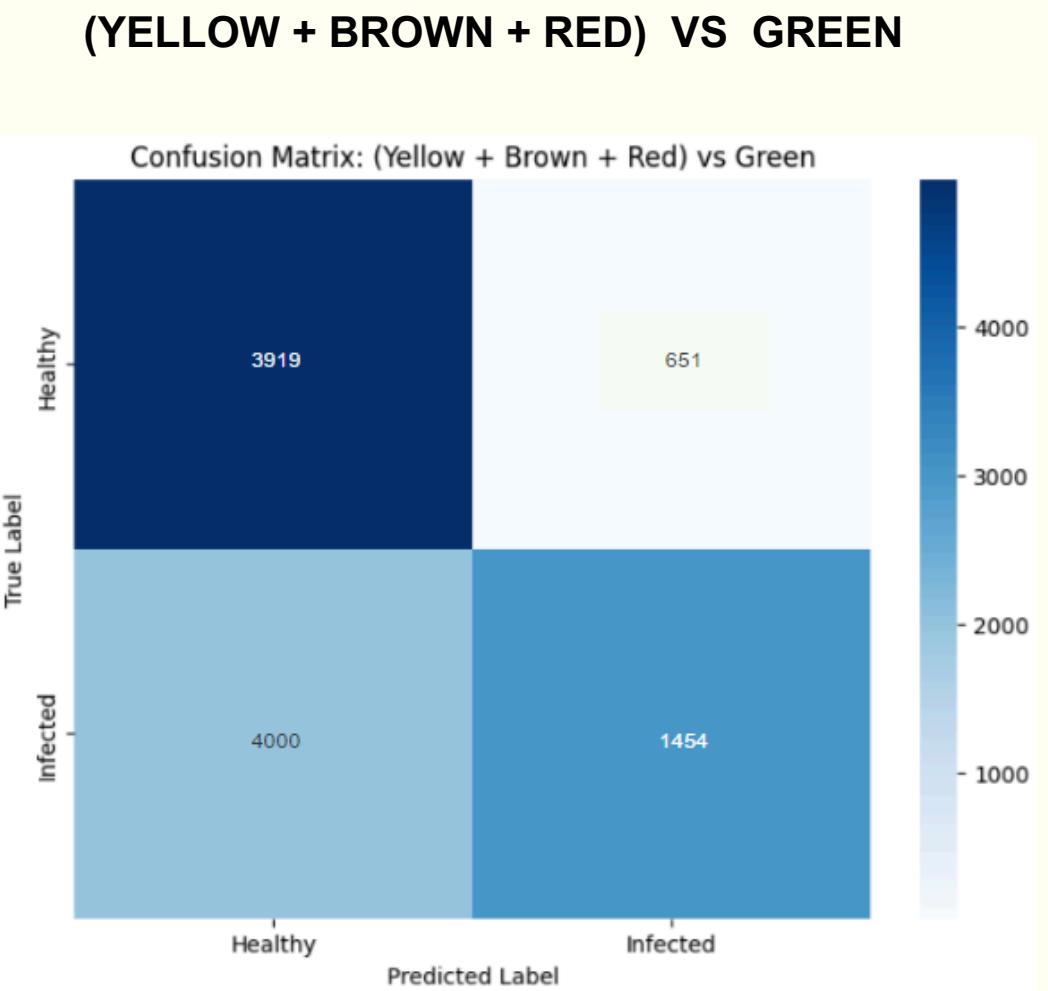
Experiment No. 1



Experiment No. 2



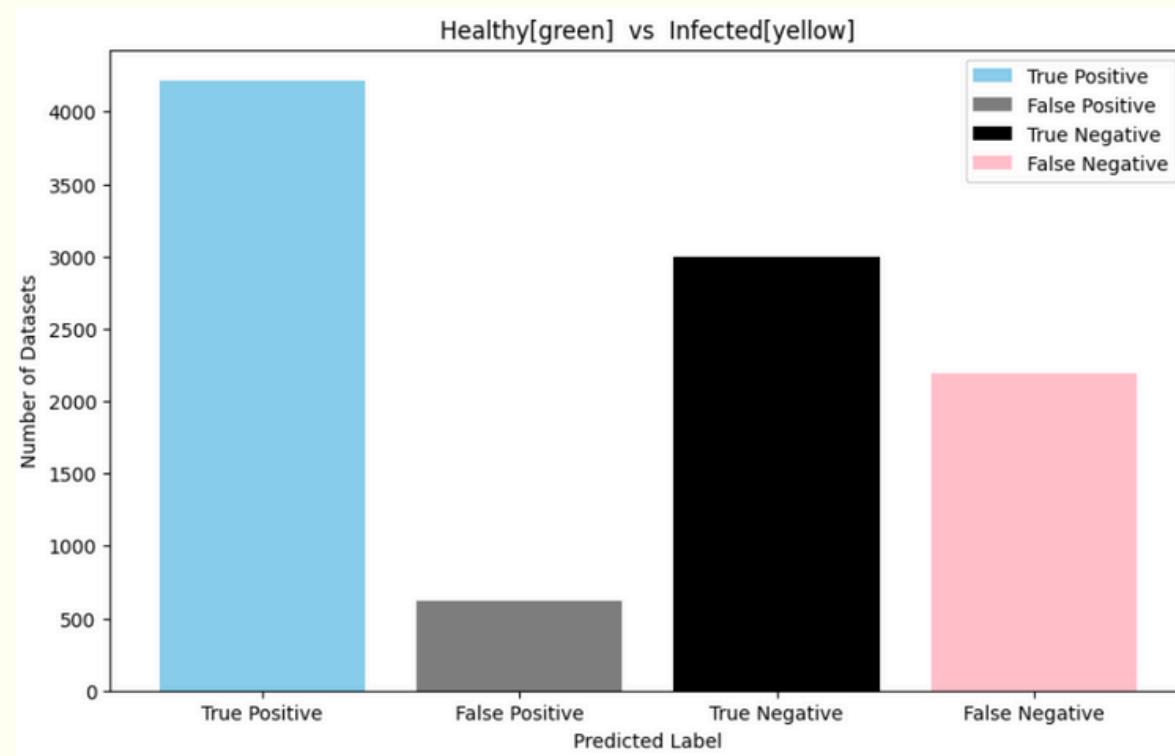
Experiment No. 3



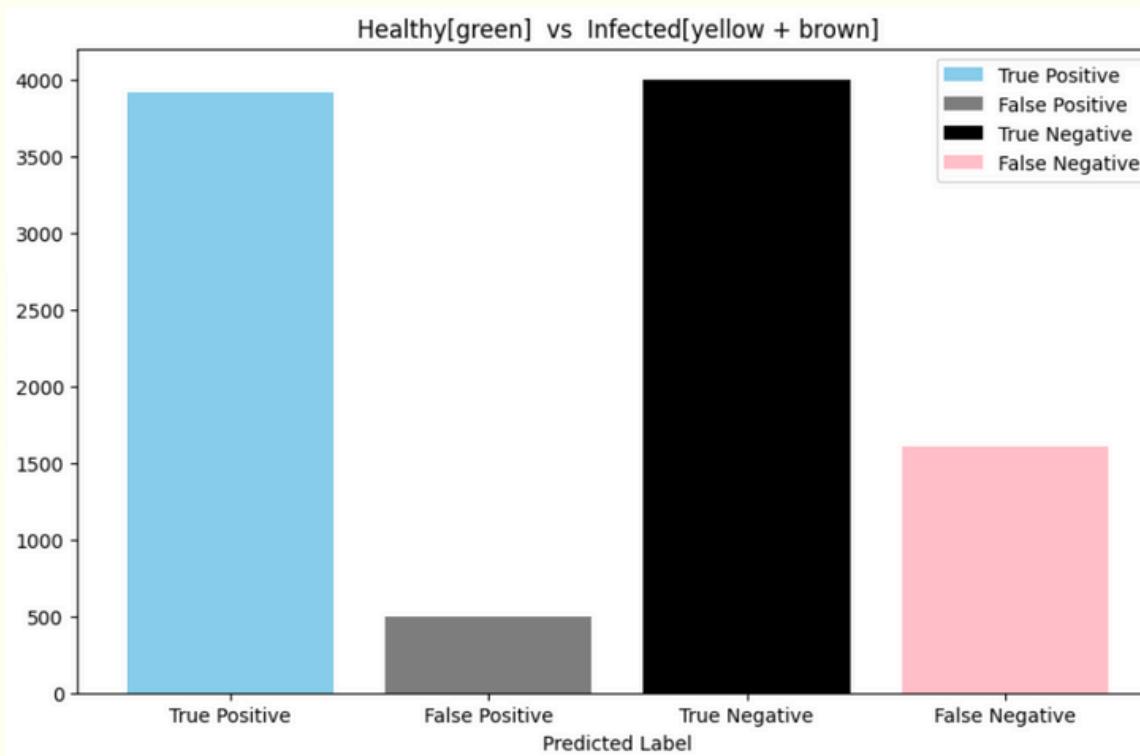
A. Customized Color - Comparasion Feature Extraction

USING RGB SPACE BAR GRAPH RESULTS

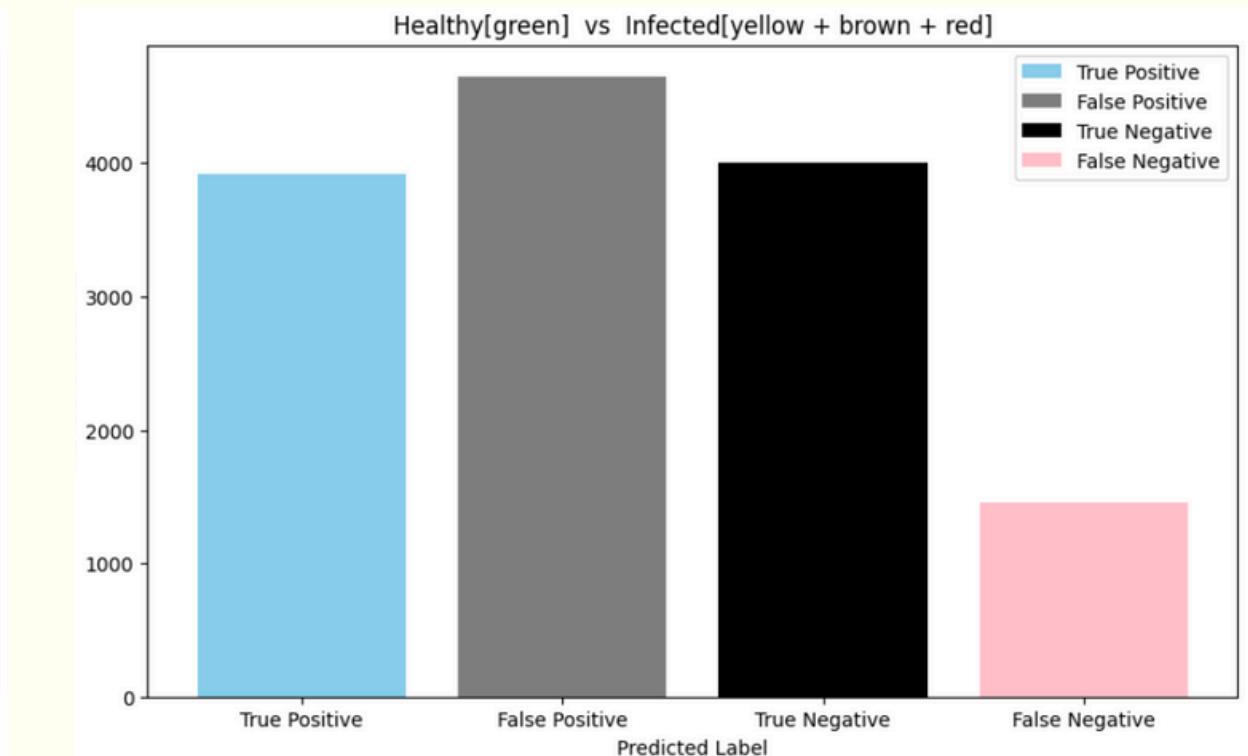
Experiment No. 1



Experiment No. 2



Experiment No. 3



ACCURACY: 72%

PRECISION: 83%

ACCURACY: 79%

PRECISION: 89%

ACCURACY: 79%

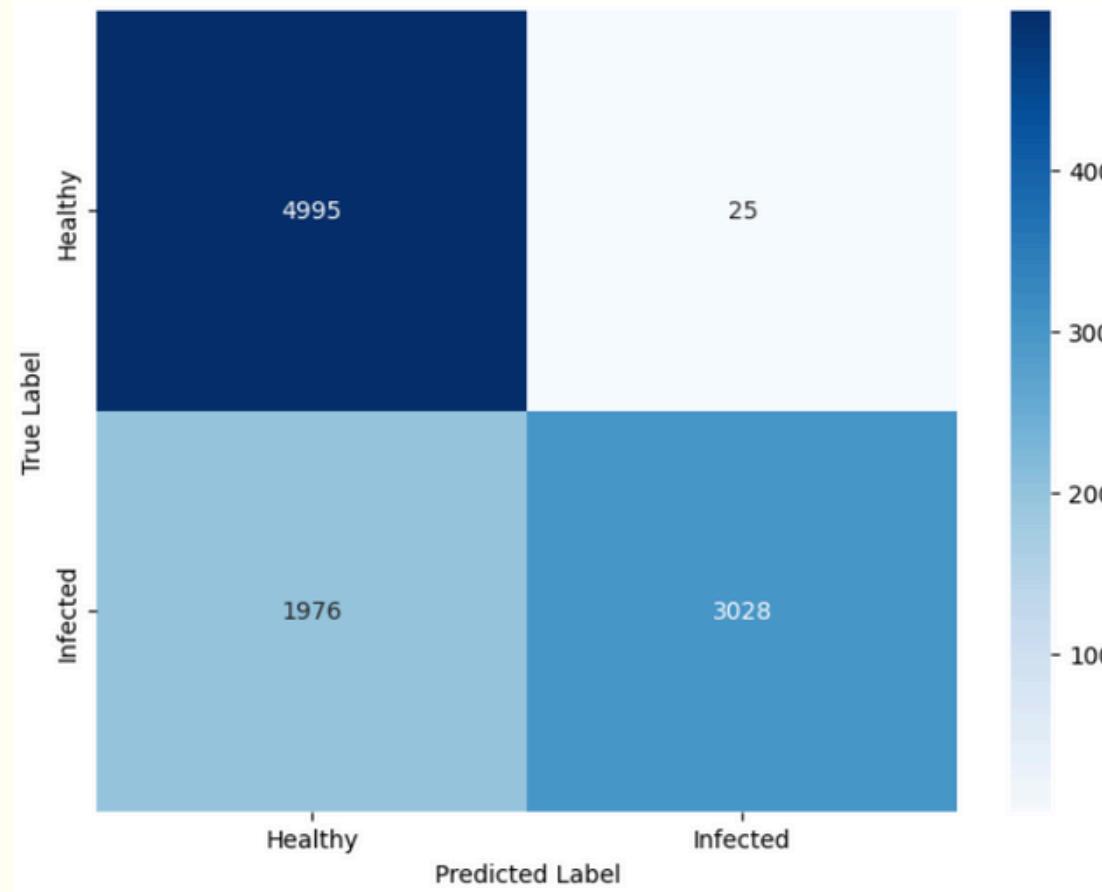
PRECISION: 86%

A. Customized Color - Comparasion Feature Extraction

USING HUE CHANNEL CONFUSION MATRIX RESULTS

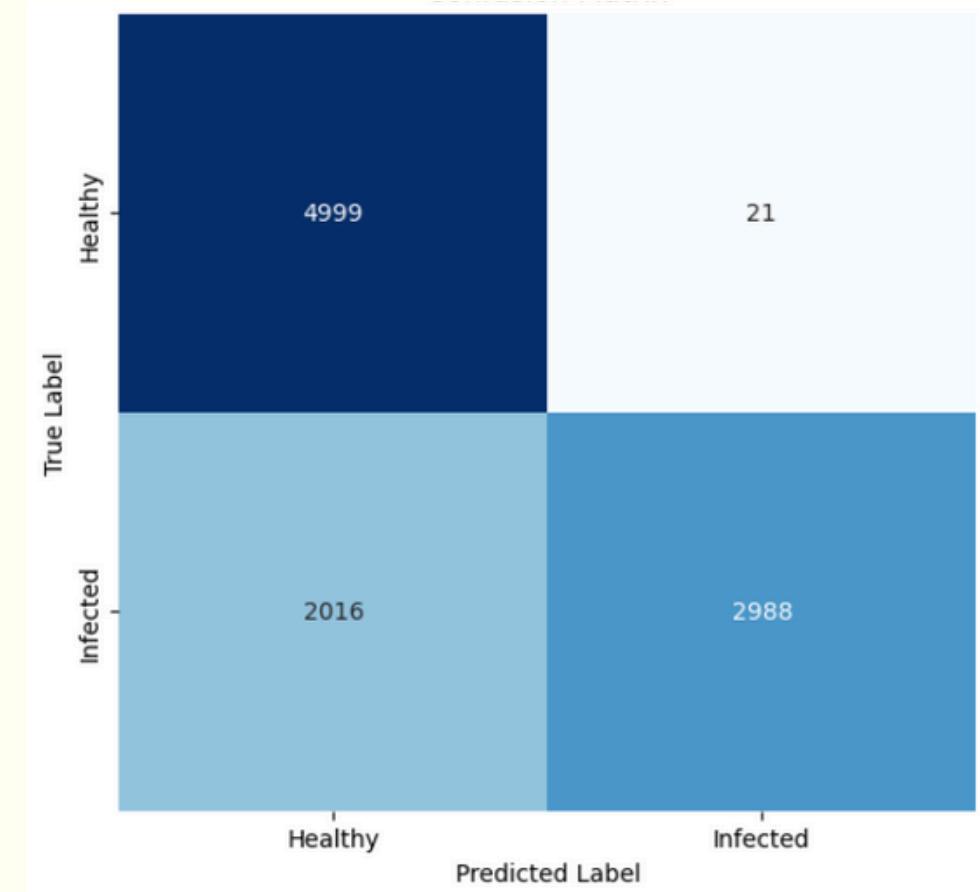
Experiment No. 1

YELLOW VS GREEN



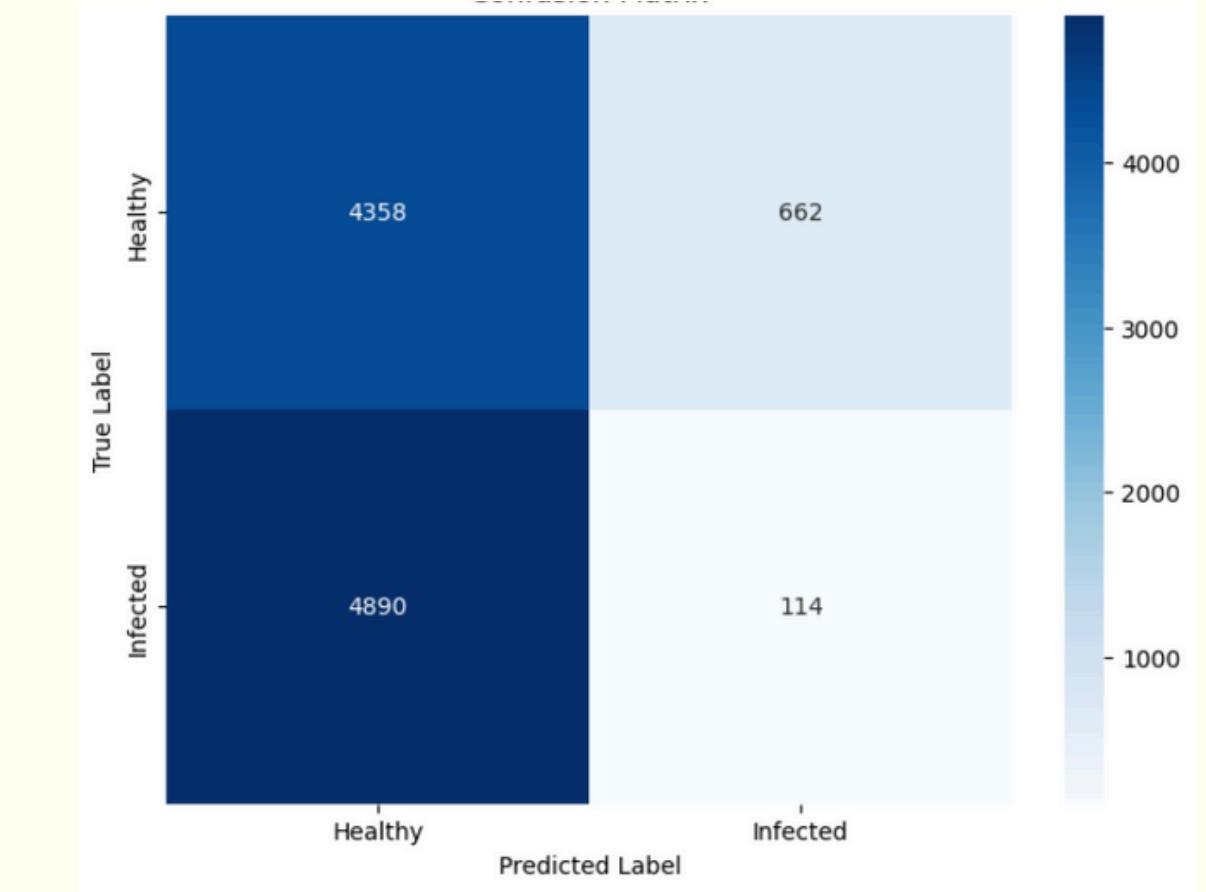
Experiment No. 2

(YELLOW + BROWN) VS GREEN



Experiment No. 3

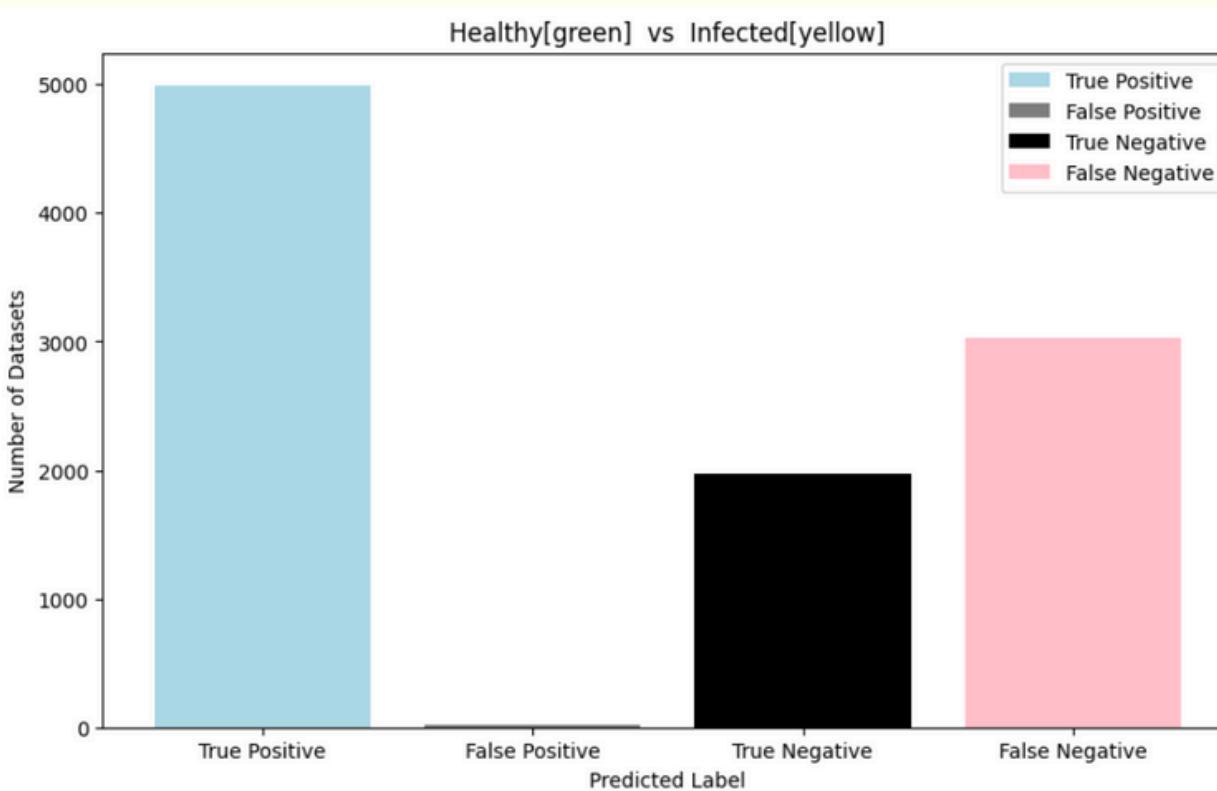
(YELLOW + BROWN + RED) VS GREEN



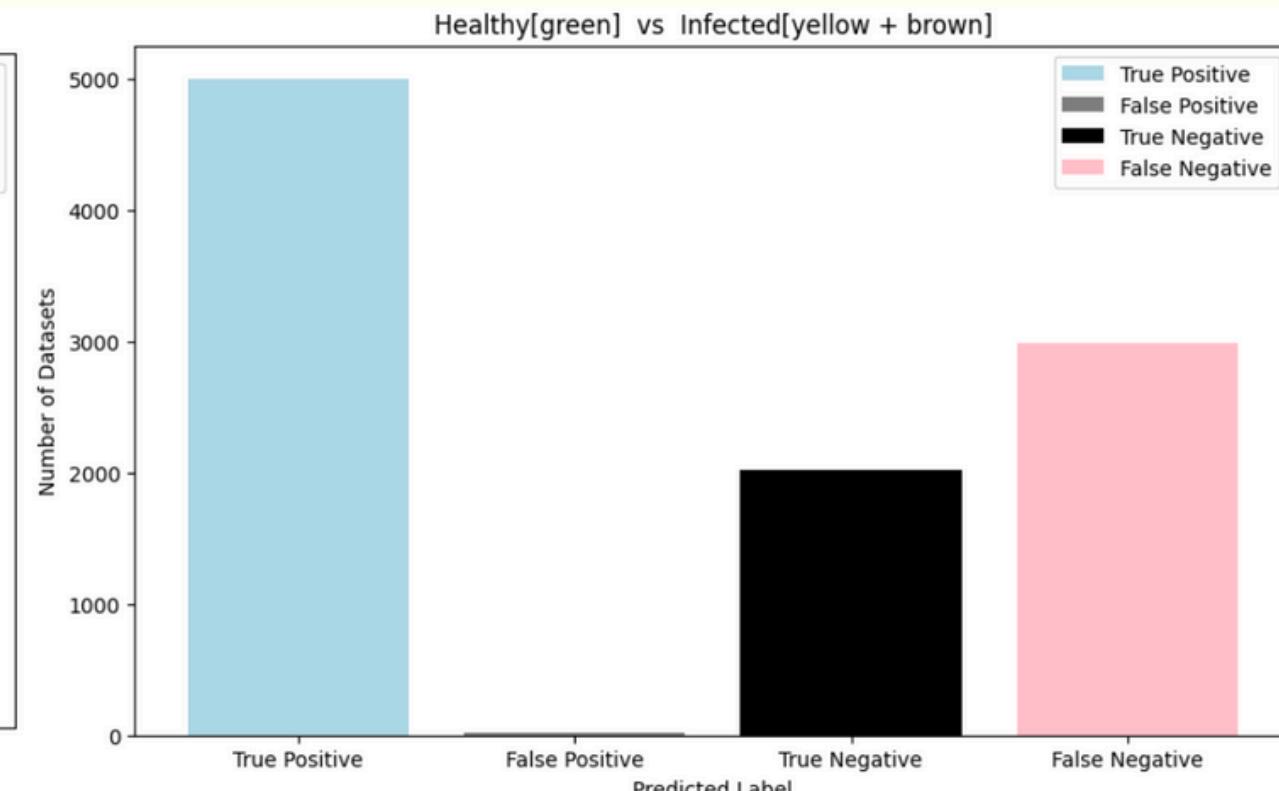
A. Customized Color - Comparasion Feature Extraction

USING HUE SPACE BAR GRAPH RESULTS

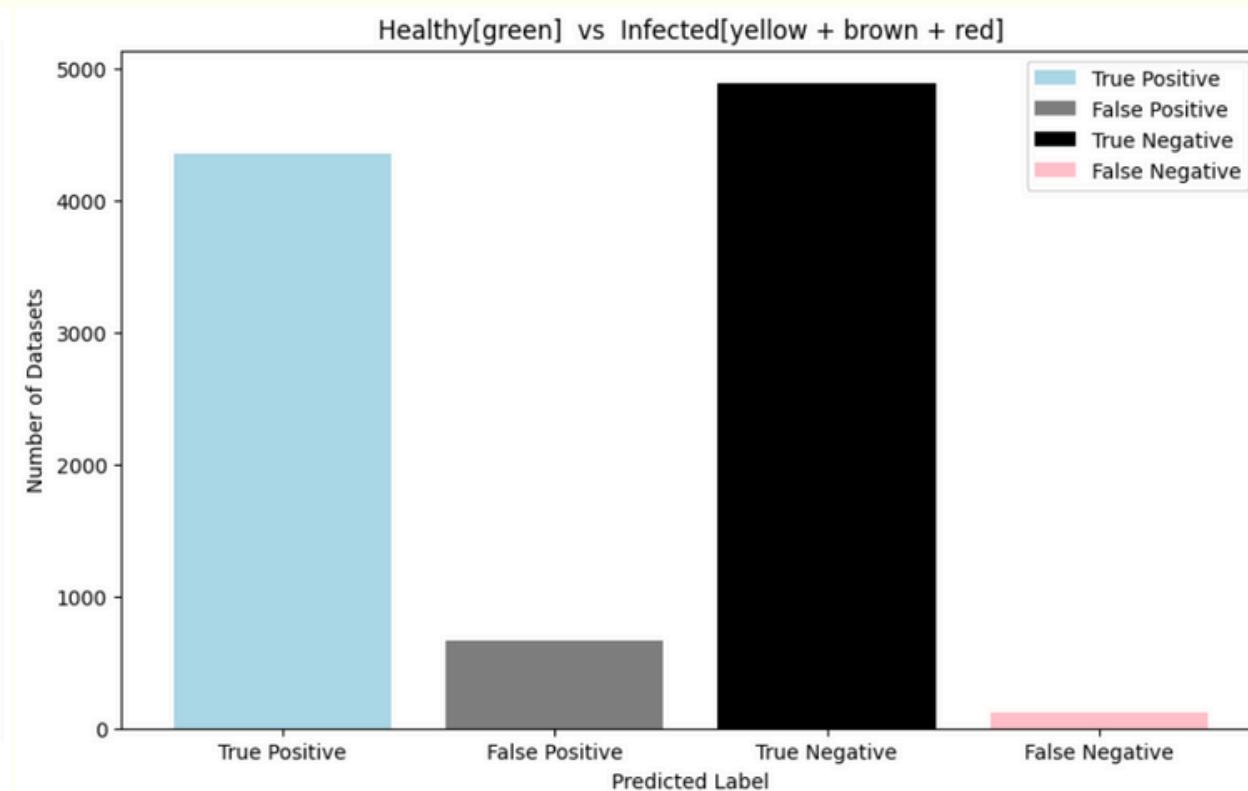
Experiment No. 1



Experiment No. 2



Experiment No. 3



ACCURACY: 69%

PRECISION: 99%

ACCURACY: 69%

PRECISION: 79%

ACCURACY: 92%

PRECISION: 86%

B.Global Color Histogram Results

USING HSV CHANNEL

In the initial experiment, the accuracy and precision were **82%** and **85%**, respectively. After refining and tuning to have a proper implementation, the updated results are as follows:

Considering 5 - Folds Cross Validation

Training 50% : Testing: 50%

```
Cross-Validation Scores: [0.95461347 0.97157107 0.96309227 0.95860349 0.9491018 ]
Mean CV Accuracy: 0.9593964191318112
```

```
Accuracy: 0.9596967278531524
Precision: 0.9899299474605955
Classification Report:
precision    recall   f1-score   support
 healthy      0.93     0.99      0.96     2572
infected      0.99     0.93      0.96     2440
accuracy
macro avg     0.96     0.96      0.96     5012
weighted avg  0.96     0.96      0.96     5012
```

Training 60% : Testing: 40%

```
Cross-Validation Scores: [0.97339983 0.97672485 0.9758936 0.98836243 0.96838602]
Mean CV Accuracy: 0.9765533476347954
```

```
Accuracy: 0.9650872817955112
Precision: 0.9896907216494846
Classification Report:
precision    recall   f1-score   support
 healthy      0.94     0.99      0.97     2065
infected      0.99     0.94      0.96     1945
accuracy
macro avg     0.97     0.96      0.96     4010
weighted avg  0.97     0.97      0.97     4010
```

Training 70% : Testing: 30%

```
Cross-Validation Scores: [0.95461347 0.97157107 0.96309227 0.95860349 0.9491018 ]
Mean CV Accuracy: 0.9593964191318112
```

```
Accuracy: 0.964095744680851
Precision: 0.990572878897752
Classification Report:
precision    recall   f1-score   support
 healthy      0.94     0.99      0.97     1547
infected      0.99     0.93      0.96     1461
accuracy
macro avg     0.97     0.96      0.96     3008
weighted avg  0.97     0.96      0.96     3008
```

ACCURACY: 95%

PRECISION: 98%

ACCURACY: 96%

PRECISION: 98%

ACCURACY: 96%

PRECISION: 99%

B.Global Color Histogram Results

USING RGB CHANNEL

Using RGB Channel the results are the following:

Considering 5 - Folds Cross Validation

Training 50% : Testing: 50%

Training 60% : Testing: 40%

Training 70% : Testing: 30%

```
Cross-Validation Scores: [0.97339983 0.97672485 0.9758936 0.98836243 0.96838602]
Mean CV Accuracy: 0.9765533476347954
```

```
Cross-Validation Scores: [0.98148148 0.97719173 0.97647897 0.98574483 0.97576622]
Mean CV Accuracy: 0.9793326469734167
```

```
Cross-Validation Scores: [0.97008973 0.96311067 0.9740519 0.97305389 0.96606786]
Mean CV Accuracy: 0.96927481029964
Accuracy: 0.981245011971269
Precision: 0.9945193929173693
Classification Report:
precision recall f1-score support
 healthy 0.97 0.99 0.98 2572
infected 0.99 0.97 0.98 2440
accuracy 0.98 0.98 0.98 5012
macro avg 0.98 0.98 0.98 5012
weighted avg 0.98 0.98 0.98 5012
```

```
Accuracy: 0.9887780548628429
Precision: 0.9953076120959332
Classification Report:
precision recall f1-score support
 healthy 0.98 1.00 0.99 2065
infected 1.00 0.98 0.99 1945
accuracy 0.99 0.99 0.99 4010
macro avg 0.99 0.99 0.99 4010
weighted avg 0.99 0.99 0.99 4010
```

```
Accuracy: 0.9900265957446809
Precision: 0.9958419958419958
Classification Report:
precision recall f1-score support
 healthy 0.98 1.00 0.99 1547
infected 1.00 0.98 0.99 1461
accuracy 0.99 0.99 0.99 3008
macro avg 0.99 0.99 0.99 3008
weighted avg 0.99 0.99 0.99 3008
```

ACCURACY: 98%

PRECISION: 99%

ACCURACY: 98%

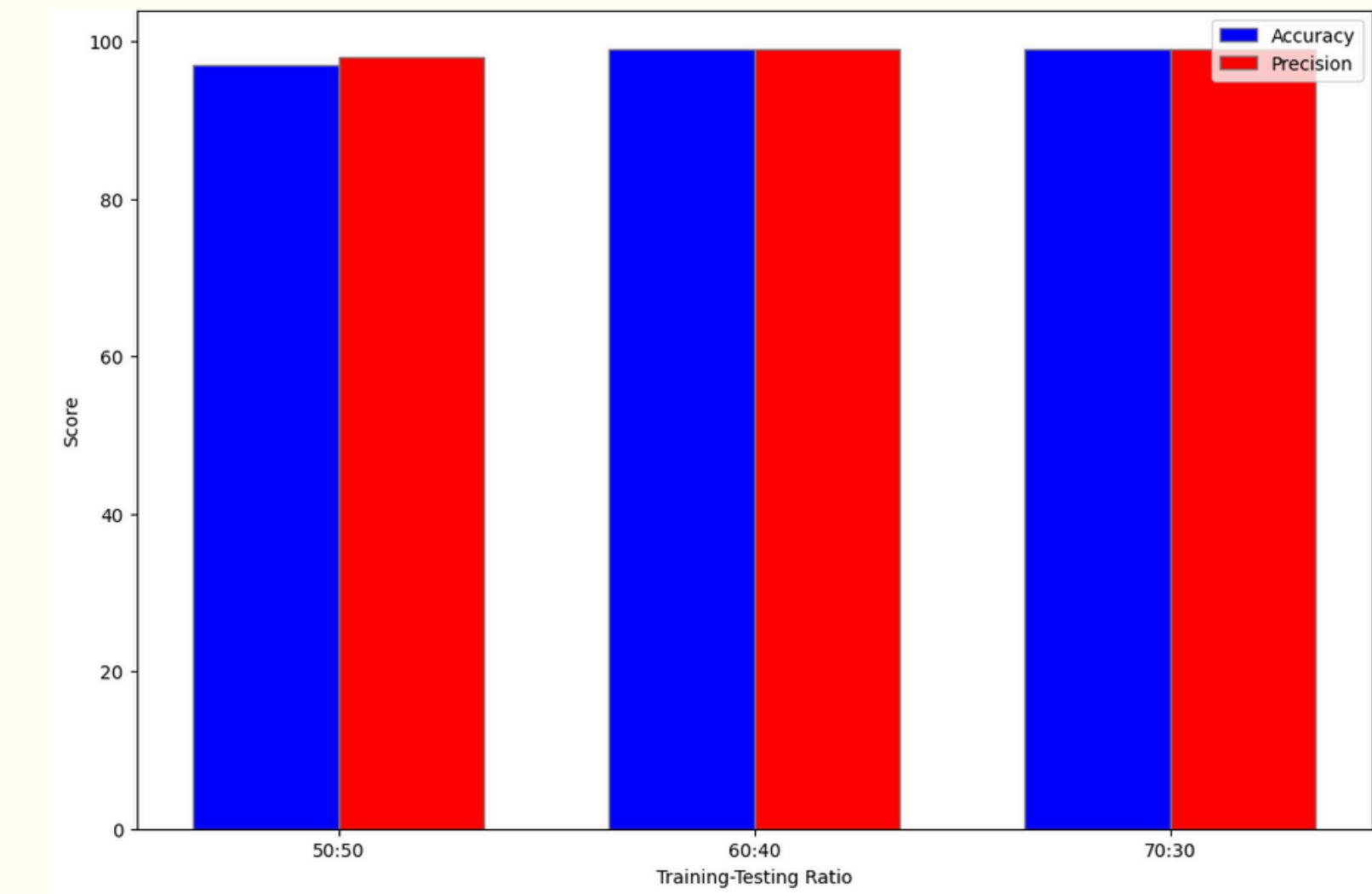
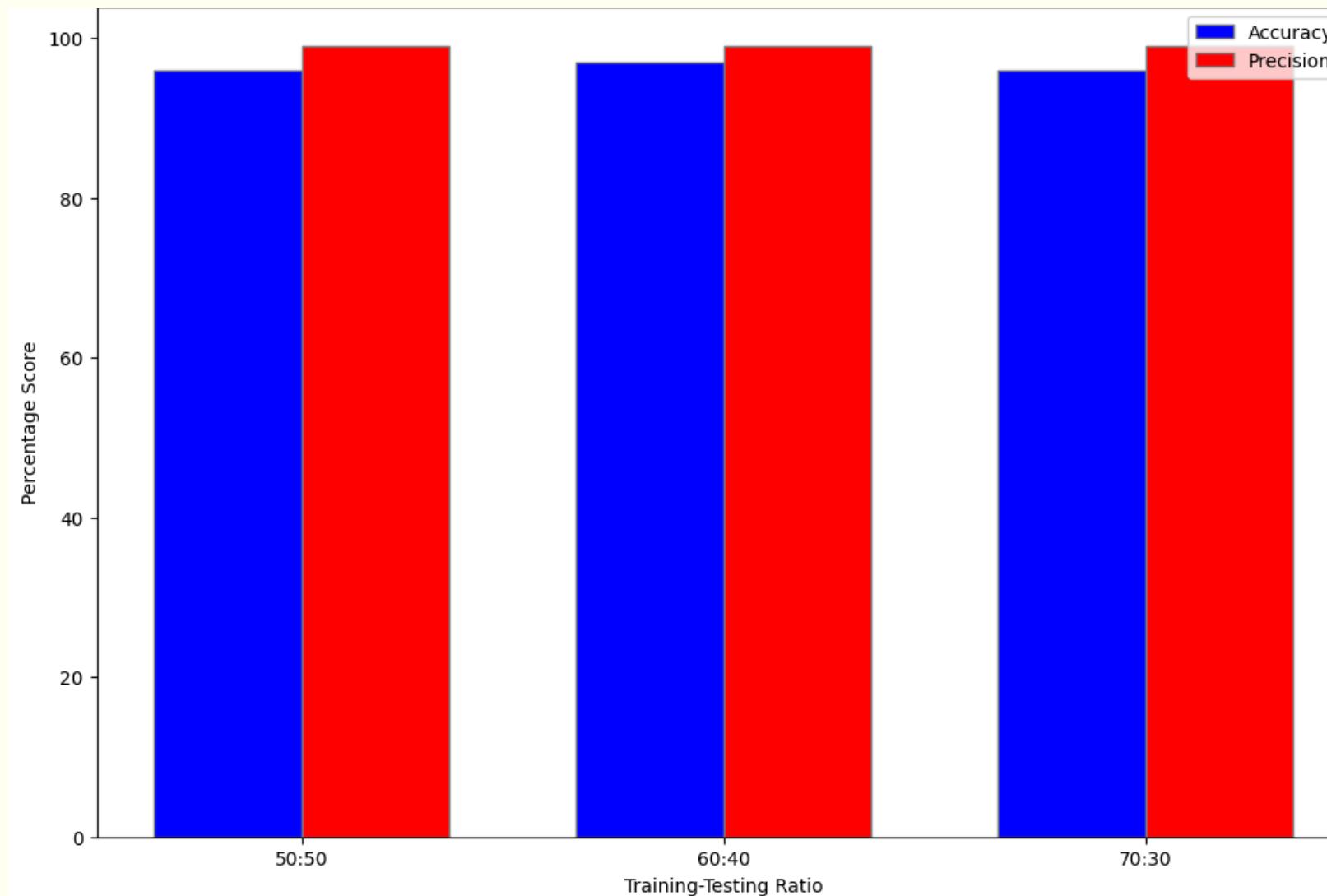
PRECISION: 99%

ACCURACY: 99%

PRECISION: 99%

B.Global Color Histogram Results

USING HSV CHANNEL RGB GRAPHS RESULT



USING HSV CHANNEL

USING RGB CHANNEL

B. Color Moments

USING HSV CHANNEL

In the initial experiment without changing to HSV channel , the accuracy and precision were **72%** and **83%**, respectively. After refining and tuning to have a proper implementation, the updated results are as follows:

Considering 5 - Folds Cross Validation

Training 50% : Testing: 50%

```
Cross Validation Scores: [0.98604187 0.99102692 0.98802395 0.98502994 0.98303393]
Mean CV Score: 0.9866313235940881
Test Accuracy: 0.9890263367916999
Test Precision: 0.9890344981614791
Classification Report:
precision    recall   f1-score   support
 healthy       0.99      0.99      0.99      2572
infected       0.99      0.99      0.99      2440
accuracy        0.99      0.99      0.99      5012
macro avg       0.99      0.99      0.99      5012
weighted avg    0.99      0.99      0.99      5012
```

Training 60% : Testing: 40%

```
Cross Validation Scores: [0.99002494 0.9833749 0.98669992 0.98919368 0.98419301]
Mean CV Score: 0.9866972889462422
Test Accuracy: 0.9902743142144639
Test Precision: 0.9902766398163663
Classification Report:
precision    recall   f1-score   support
 healthy       0.99      0.99      0.99      2065
infected       0.99      0.99      0.99      1945
accuracy        0.99      0.99      0.99      4010
macro avg       0.99      0.99      0.99      4010
weighted avg    0.99      0.99      0.99      4010
```

Training 70% : Testing: 30%

```
Cross Validation Scores: [0.98076923 0.99287242 0.98218104 0.9914469 0.98717035]
Mean CV Score: 0.9868879872800044
Test Accuracy: 0.990265957446809
Test Precision: 0.9900389577282315
Classification Report:
precision    recall   f1-score   support
 healthy       0.99      0.99      0.99      1547
infected       0.99      0.99      0.99      1461
accuracy        0.99      0.99      0.99      3008
macro avg       0.99      0.99      0.99      3008
weighted avg    0.99      0.99      0.99      3008
```

ACCURACY: 98%

PRECISION: 98%

ACCURACY: 99%

PRECISION: 99%

ACCURACY: 99%

PRECISION: 99%

B. Color Moments

USING HSV CHANNEL

Considering 5 - Folds Cross Validation

Training 50% : Testing: 30% : Validation: 20%

```
Cross Validation Scores: [0.98604187 0.99102692 0.98802395 0.98502994 0.98303393]
Mean CV Score: 0.9866313235940881
Test Accuracy: 0.9890256069171932
Test Precision: 0.9890269071014883
Classification Report:
      precision    recall   f1-score   support
  healthy       0.99       0.99       0.99      1548
infected       0.99       0.99       0.99      1459
accuracy
macro avg       0.99       0.99       0.99      3007
weighted avg     0.99       0.99       0.99      3007
```

ACCURACY: 98%

PRECISION: 98%

DISCUSSION AND ANALYSIS

A. Customized Color - Comparasion Feature Extraction

Using RGB channels, in Experiment No. 1 (comparing yellow and green), the classifier achieved an accuracy of 72% and a precision of 83%. In Experiment No. 2, which involved distinguishing between a combination of yellow and brown objects versus green ones, the classifier exhibited promising performance metrics. The experiment yielded an overall accuracy of 79%, suggesting relatively strong performance in correctly classifying the majority of samples. Moreover, the precision of 89% highlights the high proportion of correctly identified combined category samples among those classified as such. In Experiment No. 3, an overall accuracy of 79% was achieved, indicating solid performance in correctly classifying the majority of samples. Additionally, the precision of 86% underscores the high proportion of correctly identified combined category samples.

Using Hue channel, the experiment 1 yielded an accuracy of 69%, suggesting a moderate level of overall correct classifications. Additionally, the precision of 99% underscores the high proportion of correctly identified target category samples among those classified as such, indicating a high level of confidence in the classifier's positive predictions. The experiment 2 yielded an accuracy of 69%, suggesting a moderate level of overall correct classifications. Additionally, the precision of 79% highlights the proportion of correctly identified target category samples among those classified as such, indicating a reasonable level of confidence in the classifier's positive predictions. Experiment 3 achieved an overall accuracy of 79%, signaling a commendable performance in accurately categorizing the bulk of the samples. Furthermore, the precision of 86% emphasizes the substantial number of correctly identified samples within the combined category.

Based on the results, RGB channels generally offer better overall accuracy, while Hue channels excel in precise identification of specific categories.

DISCUSSION AND ANALYSIS

B. Global Color Histogram

Utilizing a Linear Support Vector Machine (SVM) as the intelligence system for object classification with the HSV channel yielded notable improvements in accuracy and precision. Initially, the system achieved 82% accuracy and 85% precision. However, after refining and tuning the implementation, significant enhancements were observed across various training-testing splits. With a 50%-50% split and 5-fold cross-validation, the accuracy surged to 95% with a precision of 98%. Even with a 60%-40% split, the accuracy reached an impressive 96%, maintaining a precision of 99%. Moreover, with a more stringent 70%-30% split, the system maintained a high accuracy of 96% while still achieving a precision of 99%.

Using the RGB channel with a training-testing split of 50%-50% and 5-fold cross-validation, the accuracy increased to an impressive 98% with a precision of 99%. Further, with a 60%-40% split, the accuracy soared to 98%, while maintaining a precision of 99%. Similarly, with a 70%-30% split, the accuracy remained at 99%, still maintaining a precision of 99%.

Comparing these results, it's evident that both approaches - SVM with HSV and SVM with RGB - led to remarkable improvements in accuracy and precision. However, the RGB channel approach showcased slightly higher accuracies and maintained comparable precision levels across different training-testing splits. This suggests that while the HSV channel may offer valuable insights into the color characteristics of objects, the RGB channel might provide more robust features for accurate classification.

DISCUSSION AND ANALYSIS

C. Color Moments

Utilizing Color Moments and a Linear Support Vector Machine (SVM) for object classification initially resulted in an accuracy of 72% and precision of 83%. Then, the significant improvements were observed with a 50%-50% training-testing split and 5-fold cross-validation, the accuracy increased dramatically to 98%, with a precision of 99%. Further enhancements were observed with a 60%-40% split, where the accuracy reached an impressive 99%, maintaining a precision of 99%. Similarly, with a more stringent 70%-30% split, the accuracy remained at 99%, still maintaining a precision of 99%. Additionally, when a 50% Training - 30% - Testing - 20% - Validation split was employed for training, testing, and validation respectively, the model maintained high accuracy and precision levels.

The model showcases robust performance across various training-testing splits and demonstrates high reliability in accurately classifying objects.

DISCUSSION AND ANALYSIS

Overall Analysis

In the overall analysis and observation of Phytoplasma disease classification, the proper selection of feature extraction and implementation lead to commendable accuracy and precision. In the study, proper image pre-processing and segmentation ensure that the datasets are clean. Then, the selection of the right feature extraction approach, which is color, is crucial. Color is a common characteristic that we observe which separates healthy from infected leaves. Finally, the proper implementation of the selected Linear Support Vector Machine also contributes to achieving high accuracy and precision.

5 . C O N C L U S I O N

In conclusion, the study has made significant strides towards addressing the problem at hand, as it has successfully developed an efficient and robust classification model. The study's objectives, which focused on implementing and analyzing the effectiveness of color features in classifying Phytoplasma disease in cassava leaves, as well as examining the utilization of a linear support vector machine , were effectively met through the results and discussions presented.

By leveraging color features and employing a linear support vector machine, the study achieved commendable accuracy and precision in distinguishing between healthy and infected. The robust classification model developed in the study provides a valuable tool for accurately identifying Phytoplasma disease symptoms based on visual characteristics captured in images. The study, alongside its implications for resource optimization, cost-effective disease management, and sustainable agriculture, underscores its significance in advancing agricultural research and practice.



STUDY WEAKNESS

- The study's weakness lies in the use of artificial light during image acquisition. Artificial light can introduce inconsistencies in color representation and may not accurately reflect the natural lighting conditions in which the plants are typically observed.

STUDY STRENGTH

- The utilization of artificial light, which ensures a controlled environment and consistent distance during image acquisition. By employing artificial light, variations in lighting conditions are minimized, thereby enhancing the reliability and reproducibility of the experimental setup. Moreover, the dataset primarily comprises images of leaves, commonly found in two states: green (healthy) and yellow (infected). This stark contrast in color values between green and yellow facilitates accurate classification, as it reduces the likelihood of mislabeling leaves based on color variations alone.
- The high accuracy and precision attained by the model demonstrate its effectiveness in reliably identifying Phytoplasma disease symptoms based on visual characteristics captured in the images.



THANK YOU