

Adulteration Detection in Spices

Identifying Black Pepper Contamination with Papaya Seeds Using Machine Learning

Jugal Alan [2024AIM1004] | **Naveen Kamalla** [2024CSM1009] | **Priyanshu Shukla** [2024CSM1016]



Table of contents

01

Literature Review

02

Dataset Collection

03

Experimentation

04

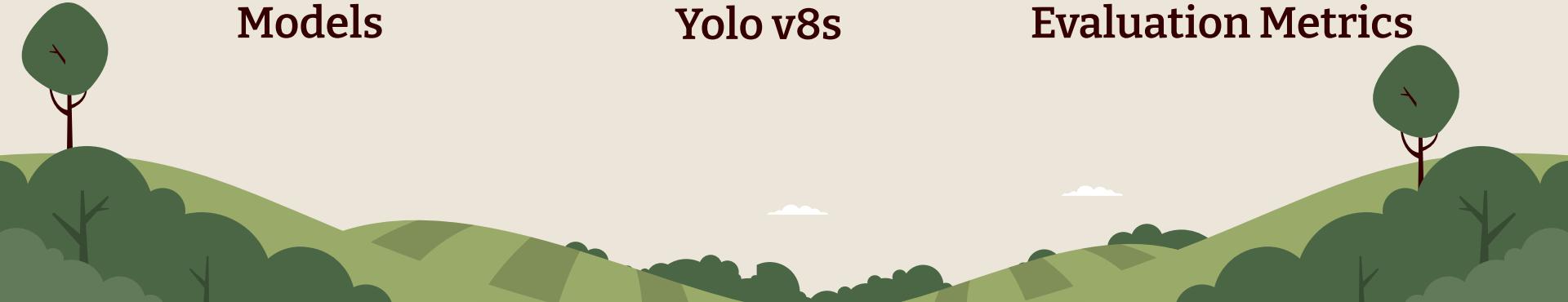
Models

05

Yolo v8s

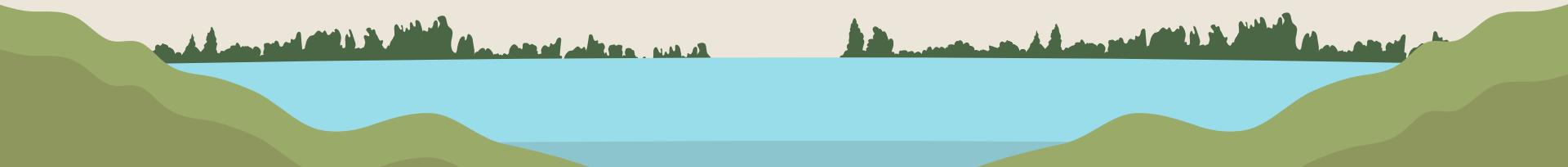
06

Evaluation Metrics



Objective

The primary objective is to create a robust machine learning model that can accurately identify and quantify the presence of papaya seeds in black pepper samples using image-based analysis. This involves classifying individual seeds and determining the level of adulteration, providing a tool for quality assurance in the spice industry.



01

Literature Review

Reviewing existing research and identifying methodologies

1. Classification of Pepper Seeds by Machine Learning Using Color Filter Array Images

This paper classifies Penja pepper seeds using color filter array (CFA) images and machine learning, achieving 87% accuracy with an SVM classifier, 80% with KNN, 79% with SGD and 83% with RF.

- Focuses on Penja pepper seed classification using CFA(Color Filter Array) images.
- Collected 5618 samples of black and white pepper seeds - 1335 PBP, 1416 PBA, 1437 PNP & 1430 PNA.
- Extracted 18 image features, including shape, texture, and color attributes.

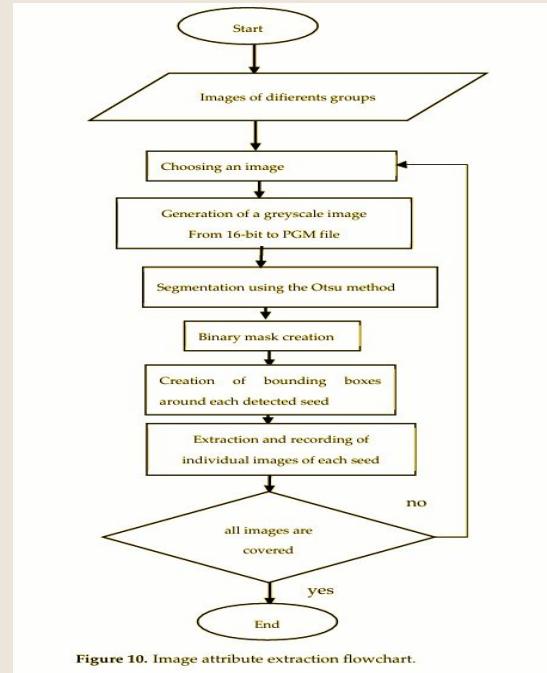


Figure 10. Image attribute extraction flowchart.

2. Adulteration Identification of Papaya Seeds in Black Pepper Using Digital Image Processing

This paper proposes a digital image processing approach to detect papaya seed adulteration in black pepper using feature extraction and k-NN classification, achieving 90% accuracy.

- Captures images of black pepper and papaya seeds under controlled conditions..
- Applies median filtering for noise reduction and extracts texture features using GLCM.
- Implemented simple KNN on obtained feature vectors.

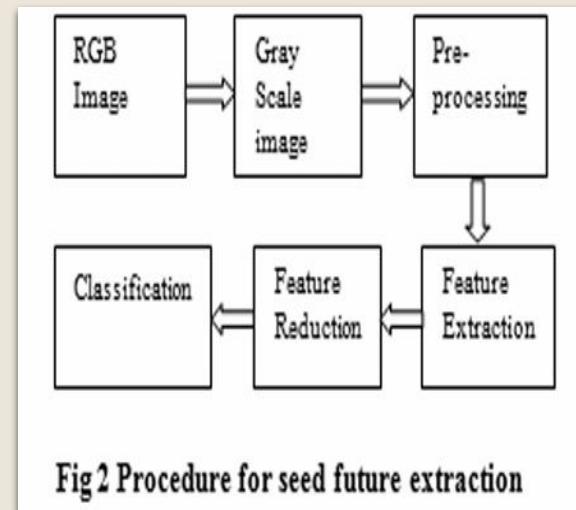
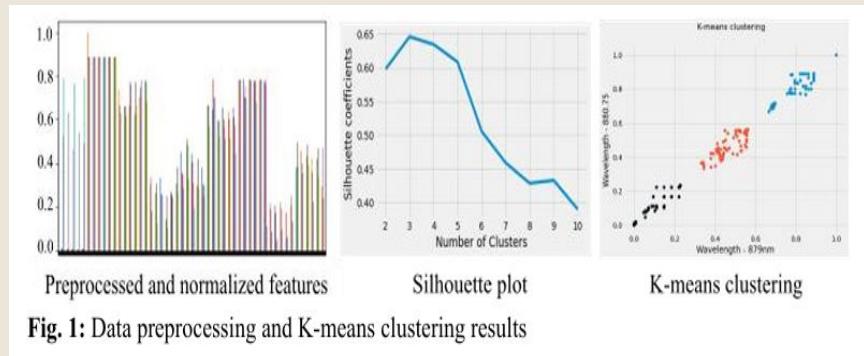


Fig 2 Procedure for seed future extraction

3. Prediction of Carica Papaya Seed Adulteration in Black Pepper using NIR Spectroscopy and Chemometric Analysis

This study employs near-infrared (NIR) spectroscopy and chemometric analysis to detect and quantify papaya seed adulteration in black pepper, achieving high prediction accuracy using Random Forest regression.

- The NIR spectral data were collected using a StellarNet Dwarf Star miniature spectrometer.
- The data was then scaled using quantile transformation and analyzed to observe the degree of separability in the dataset.
- Random Forest (RF) model demonstrated the best accuracy for predicting adulteration levels.



02

Dataset Collection

Gather images ensuring diversity

Black pepper image data



- Captured using iphone 16 camera to capture these images (48MP Fusion: 26 mm, $f/1.6$ aperture, sensor-shift optical image stabilisation, 100% Focus Pixels, support for super-high-resolution photos (24MP and 48MP))

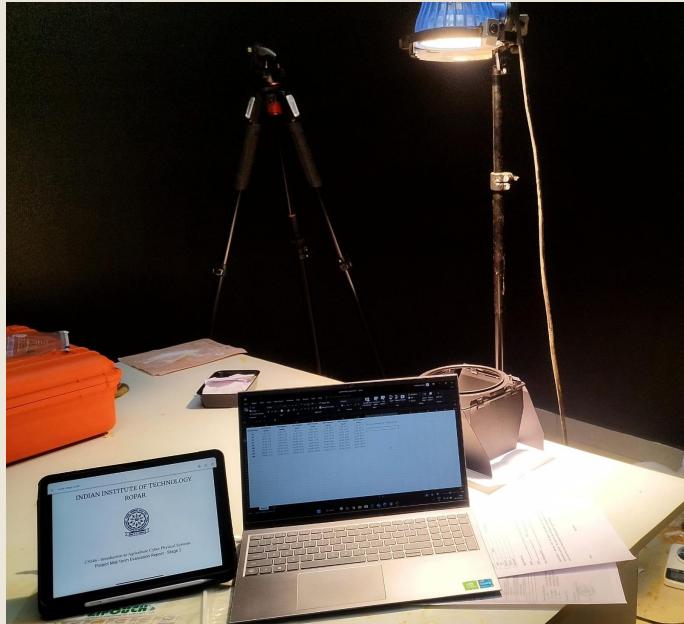
Papaya seed image data



- Captured using iphone 16 camera to capture these images (48MP Fusion: 26 mm, $f/1.6$ aperture, sensor-shift optical image stabilisation, 100% Focus Pixels, support for super-high-resolution photos (24MP and 48MP))

Adulterated data collection

The Set-up



Adulterated data collection

The Adulteration

Sample Size	5% Papaya Seeds	10% Papaya Seeds	20% Papaya Seeds	40% Papaya Seeds	50% Papaya Seeds	60% Papaya Seeds	80% Papaya Seeds	
10	9 BP, 1 PS	9 BP, 1 PS	8 BP, 2 PS	6 BP, 4 PS	5 BP, 5 PS	4 BP, 6 PS	2 BP, 8 PS	BP = Black Pepper, PS = Papaya Seeds
20	19 BP, 1 PS	18 BP, 2 PS	16 BP, 4 PS	12 BP, 8 PS	10 BP, 10 PS	8 BP, 12 PS	4 BP, 16 PS	
40	38 BP, 2 PS	36 BP, 4 PS	32 BP, 8 PS	24 BP, 16 PS	20 BP, 20 PS	16 BP, 24 PS	8 BP, 32 PS	
60	57 BP, 3 PS	54 BP, 6 PS	48 BP, 12 PS	36 BP, 24 PS	30 BP, 30 PS	24 BP, 36 PS	12 BP, 48 PS	
80	76 BP, 4 PS	72 BP, 8 PS	64 BP, 16 PS	48 BP, 32 PS	40 BP, 40 PS	32 BP, 48 PS	16 BP, 64 PS	
100	95 BP, 5 PS	90 BP, 10 PS	80 BP, 20 PS	60 BP, 40 PS	50 BP, 50 PS	40 BP, 60 PS	20 BP, 80 PS	
150	143 BP, 7 PS	135 BP, 15 PS	120 BP, 30 PS	90 BP, 60 PS	75 BP, 75 PS	60 BP, 90 PS	30 BP, 120 PS	
200	190 BP, 10 PS	180 BP, 20 PS	160 BP, 40 PS	120 BP, 80 PS	100 BP, 100 PS	80 BP, 120 PS	40 BP, 160 PS	
250	238 BP, 12 PS	225 BP, 25 PS	200 BP, 50 PS	150 BP, 100 PS	125 BP, 125 PS	100 BP, 150 PS	50 BP, 200 PS	

There are $9 \times 7 = 63$ cells that means we created 63 variations of adulteration.
For each variation we took 10 pictures for different spreads on the surface.

Adulterated data collection

The result



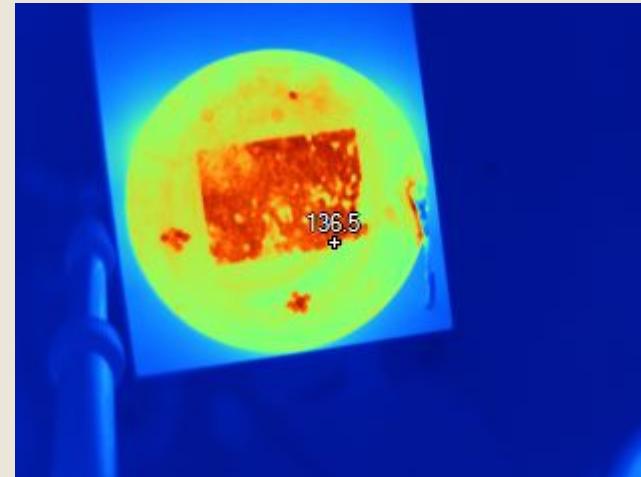
- We have used iPhone 16 camera to capture these images (48MP Fusion: 26 mm, f/1.6 aperture, sensor-shift optical image stabilisation, 100% Focus Pixels, support for super-high-resolution photos (24MP and 48MP))
- There are $9 \times 7 = 63$ cells that means we created 63 variations of adulteration.
- For each variation we took 10 pictures for different spreads on the surface.
- The total number of images we took is 630 and we're planning to create more samples using augmentation methods.

03

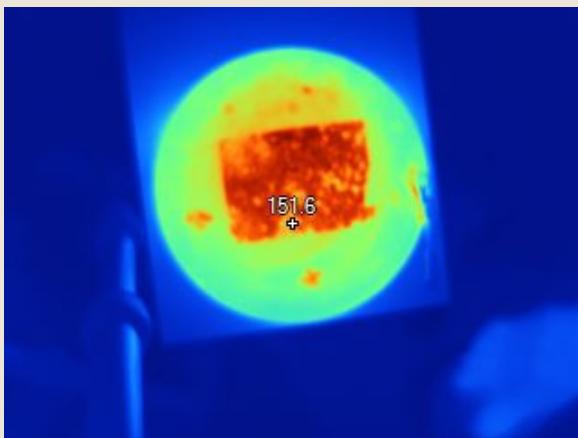
Experimentation

Experimenting with thermal images

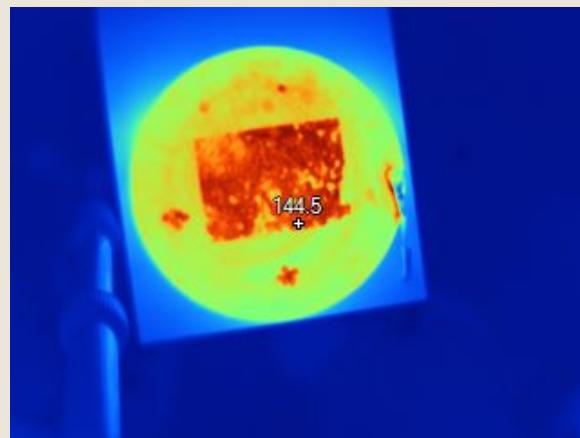
We heated some samples from papaya and pepper at 70 degree for 15 minutes to find if there was any difference in thermal scans of dried papaya seeds and black pepper. Unfortunately that didn't work out for us but here are some samples from that experiment.



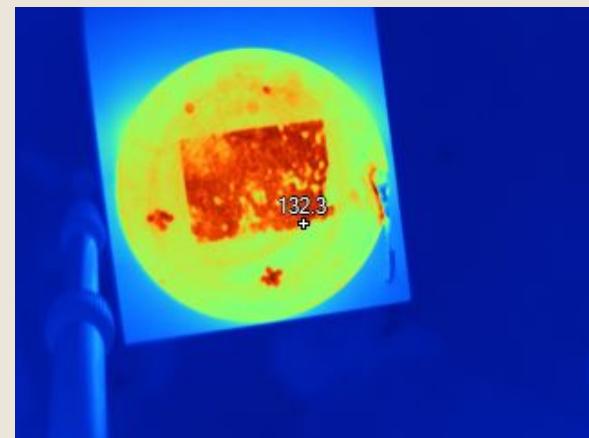
Experimenting with thermal images



00:00:00



00:07:00

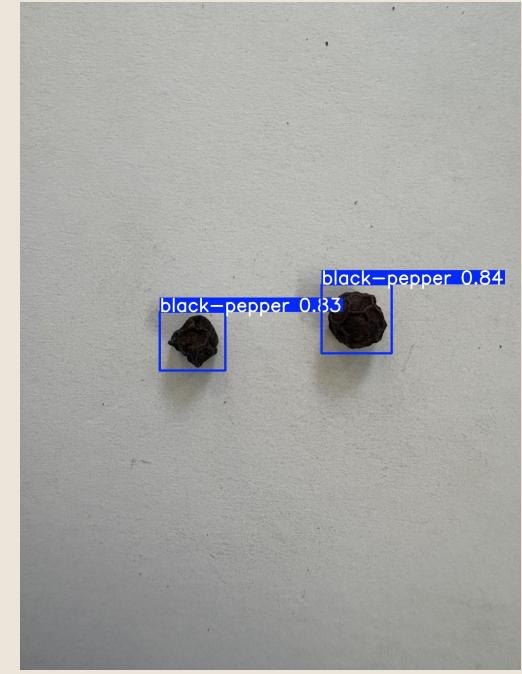


00:15:00

04

Model Experimentation

YOLO output

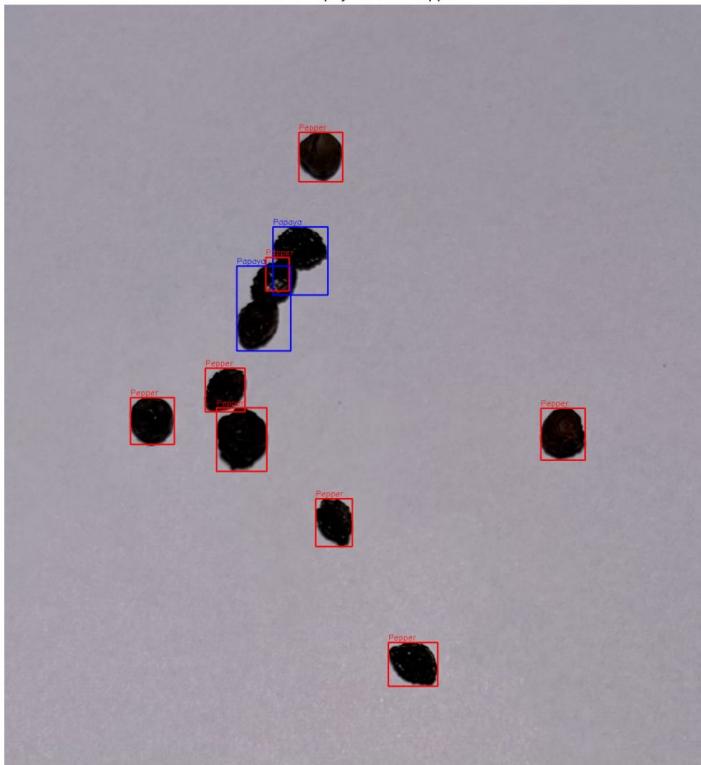


YOLO output

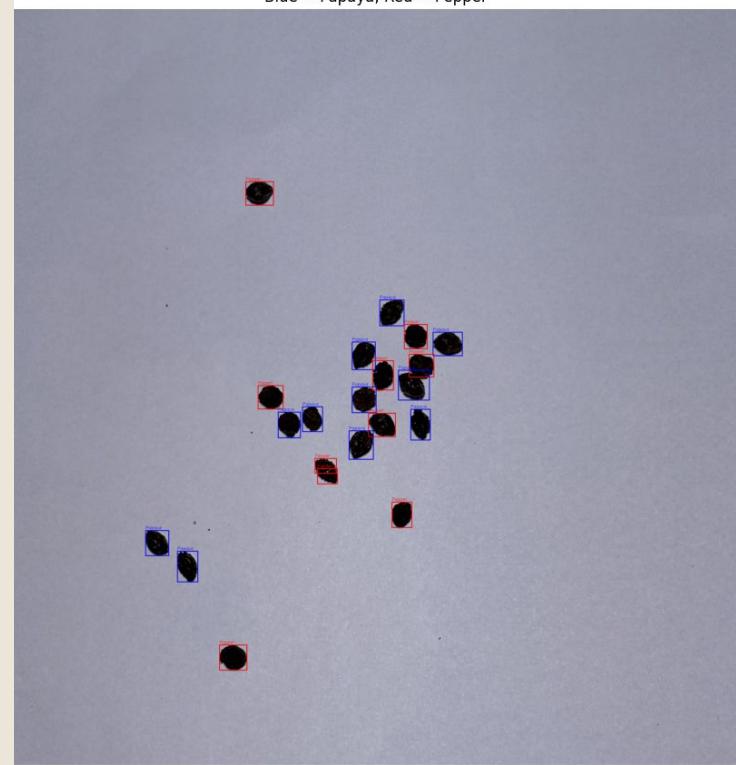


Unsupervised Learning

KMeans (Watershed) for image 10_60_02 — Total Seeds: 10 | Papaya: 2 | Pepper: 8
Blue = Papaya, Red = Pepper



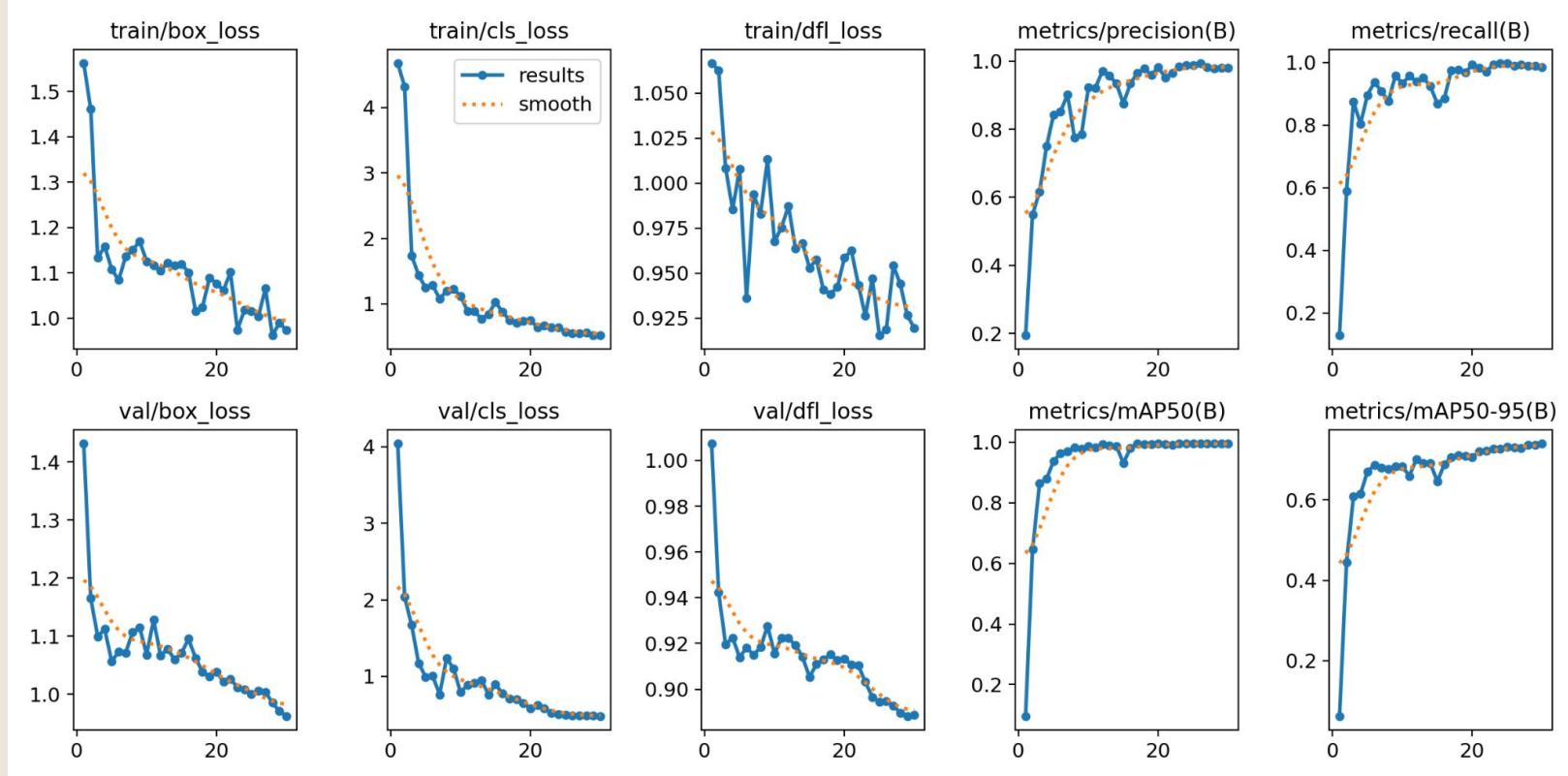
GMM Result for 20_60_07 — Total Seeds: 21 | Papaya: 11 | Pepper: 10
Blue = Papaya, Red = Pepper



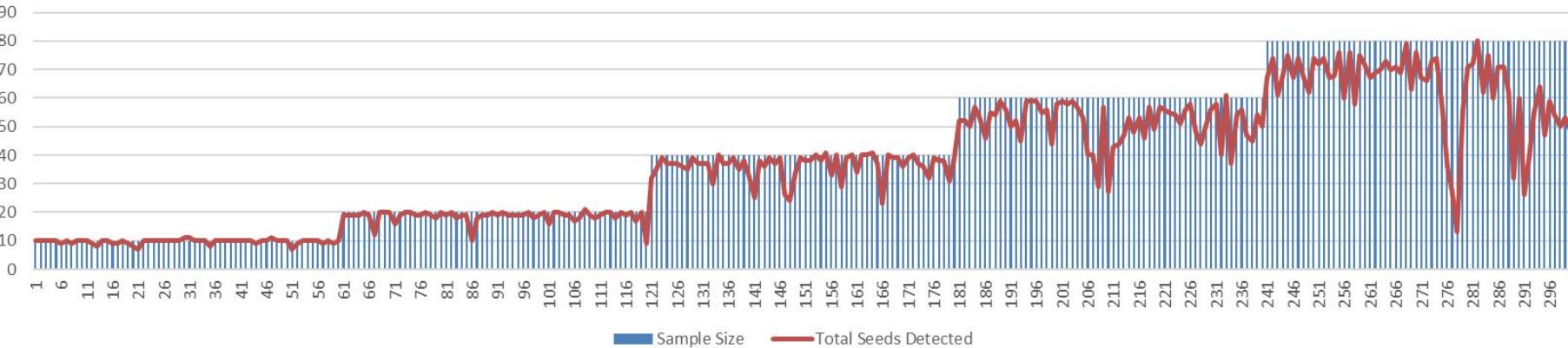
05

Yolo v8s

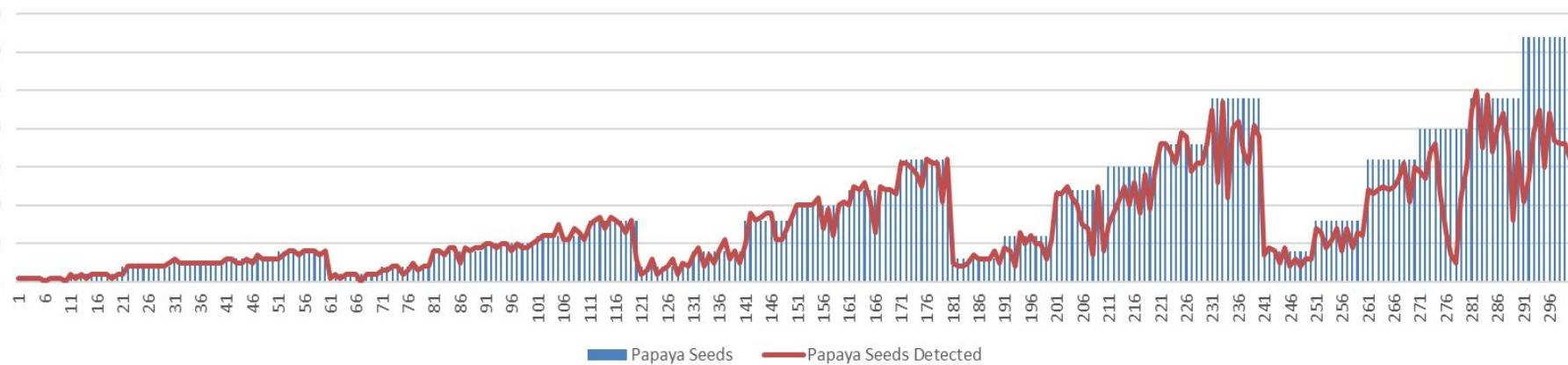
Yolo learning curve



Total Number of Seeds

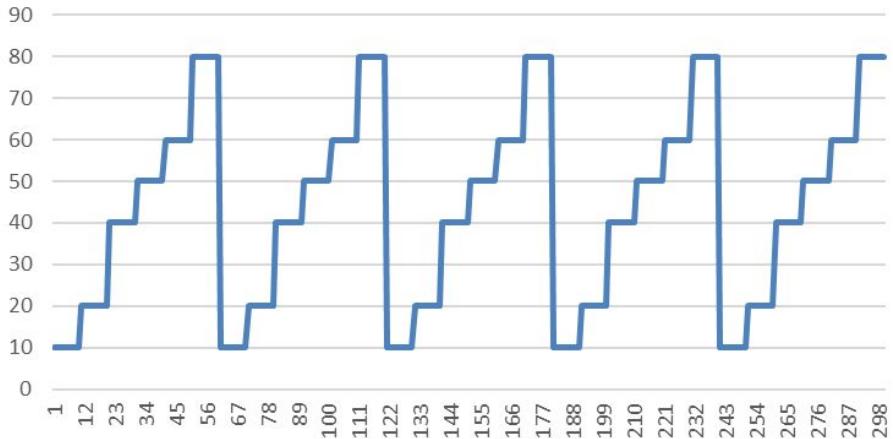


Number of papaya seeds

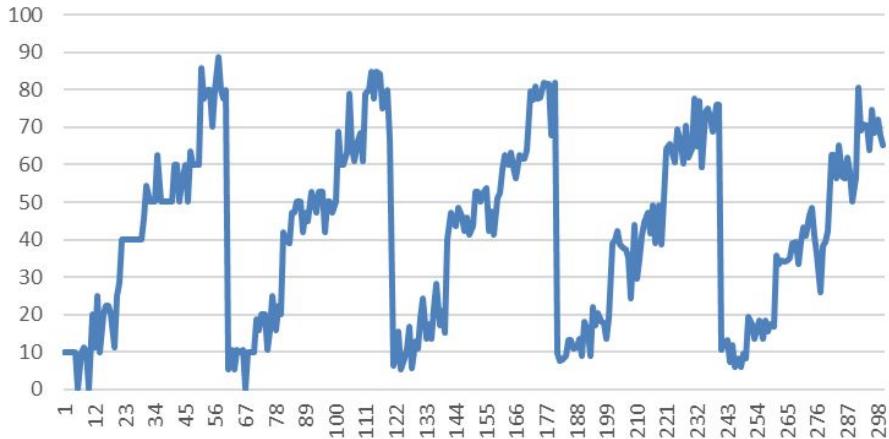


Adulteration Prediction

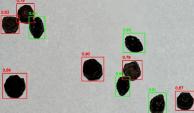
Adulteration %



Predicted Adulteration %

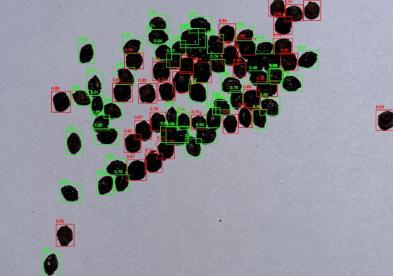


Spatial Locality



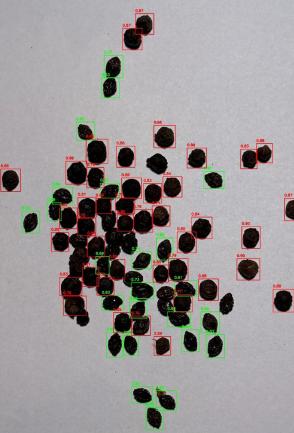
10-40-02

40-40-02

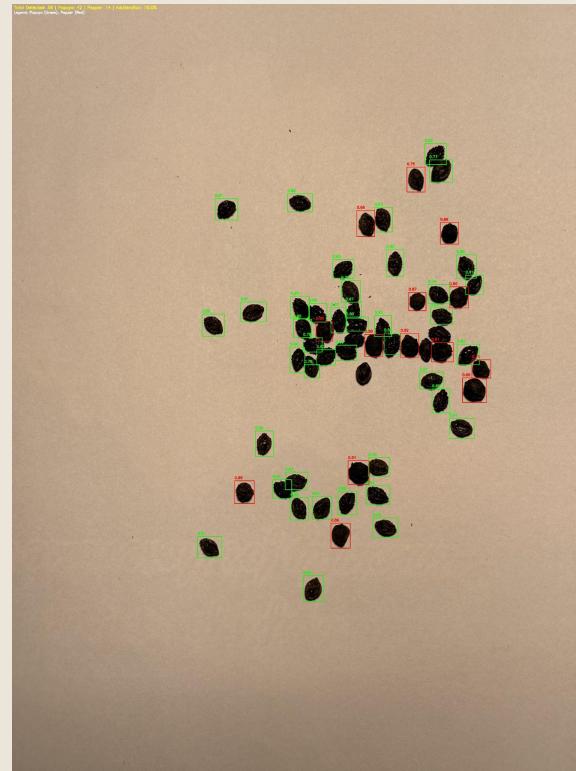


80-60-08

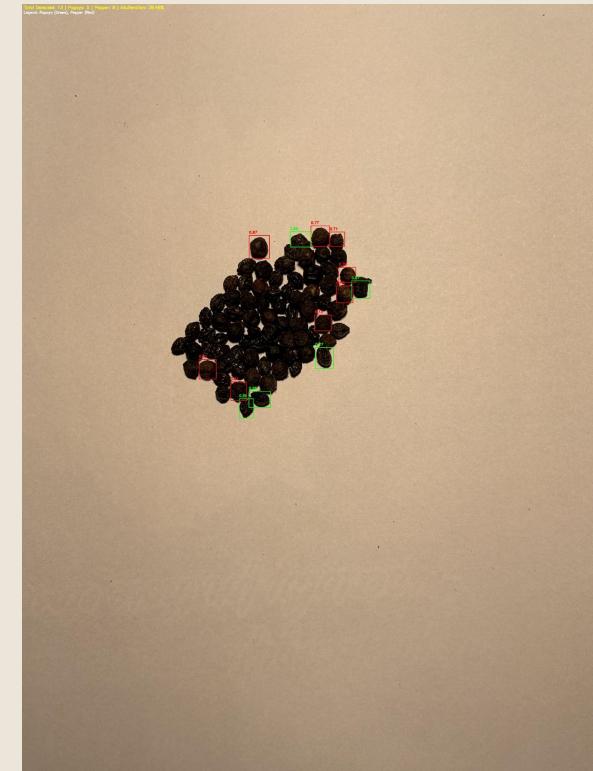
Spatial Locality



80-40-01



60-80-06



80-50-08

06

Evaluation Metrics

Mean Relative Adulteration Error (MRAE %)

- Measures the average deviation between predicted and actual adulteration levels, normalized by the actual values.
- **Formula:**

$$\text{MRAE}(\%) = \frac{|\text{Predicted} - \text{Actual}|}{\text{Actual}} \times 100$$

- **Result:**

MRAE = 12.27%

Mean Seed Detection Accuracy (MSDA %)

- Indicates the accuracy of seed type detection by comparing correctly identified seed instances to the total actual seeds.
- **Formula:**

$$\text{MSDA}(\%) = \frac{\text{Correctly Detected Seeds}}{\text{Total Seeds}} \times 100$$

- **Result:**
MSDA = 88.99%

Normalized Root Mean Square Error (NRMSE)

- Quantifies the error between predicted and actual adulteration values, normalized by the value range.
- **Formula:**

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}}{\max(y) - \min(y)}$$

- **Result:**

NRMSE = 0.0813

Thanks!

Do you have any questions?

~ Presented by

- Jugal Alan [2024AIM1004]
- Naveen Kamalla [2024CSM1009]
- Priyanshu Shukla [2024CSM1016]

