Grading and Sorting of Fruits and Vegetables using Classical Image Processing Techniques

Sreeja Reddy Yeluru
Jarvis College of Computing and Digital Media
Depaul University
Chicago, Illinois, United States
syeluru@depaul.edu

***Abstract*—** Grading and sorting fruits and vegetables are essential tasks in agriculture and food production to uphold quality standards. These activities have depended on manual inspection. With manual inspection it is often marked by inefficiency and inconsistency thereby resulting in classification errors and economic losses. This study introduces an automated grading system that employs classical image processing methods to categorize produce as good or bad. The system assesses color using HSV thresholding, examines structure through edge detection, evaluates texture with entropy analysis, and extracts features via histograms. We obtained datasets from Kaggle, cleaned them by removing irrelevant images, and ensured uniform labeling. The method involves eliminating backgrounds, refining object outlines, and classifying based on defined characteristics. Testing revealed dependable results under consistent conditions, though difficulties surfaced with fluctuating light, overlapping objects, and low-resolution images. This work highlights what classical image processing can achieve in grading produce, its shortcomings, and guiding toward future improvements. Adjustable segmentation or blending in modern learning approaches are few directions for future.

***Keywords*—** *Image processing, Fruit grading, Feature extraction, HSV segmentation, Classical techniques.*

I. Introduction

Automating the grading and sorting of fruits and vegetables plays a crucial role in cutting waste, ensuring uniform quality, and streamlining supply chains. Many farms and markets still turn to manual inspection, but this approach varies widely depending on the individual doing the work, making it unreliable. Deep learning has shown impressive accuracy in assessing quality, yet its need for vast labeled datasets, powerful computing setups, and specialized tools puts it out of reach for smaller operations.

This research explores classical image processing techniques as an efficient and cost-effective alternative. The goal is to classify produce using handcrafted features such as:

- HSV histograms to study color differences
- Sobel edge density to measure surface texture
- Shannon entropy to gauge texture complexity

Unlike systems built on machine learning, this does not demand heavy computing resources or lengthy training, offering a solution that fits well for small- to medium-sized agricultural businesses. Our approach achieved classification accuracy of approximately 63–65% across training, validation, and test datasets, highlighting both its effectiveness and challenges.

II. Background

Efforts to grade fruit and vegetable quality using images have long drawn on both straightforward image processing and more advanced machine learning techniques.

- Reference [1] crafted a fuzzy logical system to evaluate potatoes based on their texture and shape.
- Reference [2] worked with classical image processing to classify fruits, but found it tough to apply their method to a wide range of produce.
- Reference [3] laid groundwork with edge and color segmentation strategies. They noted that changing light conditions often caused problems.

Most current solutions either rely on resource heavy models or focus narrowly on specific types of produce. In this study we aimed for wide range of produce using well established image processing methods to tackle grading across various fruits and vegetables.

III. Methodology

We designed a step-by-step image processing approach for this study and made sure that covers image preparation, object isolation, feature analysis, and classification. The full process is outlined in the below “Fig. 1”.

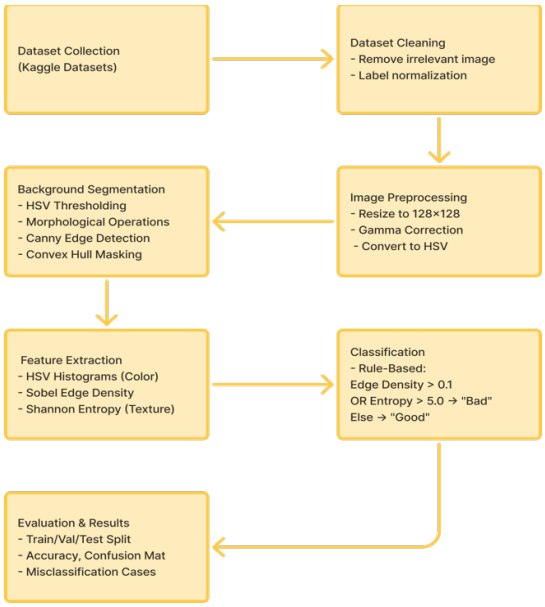

Fig. 1. Image Processing Pipeline.

A. Dataset Preparation

To build and test our system, we gathered datasets from Kaggle:

- *Mango Dataset:* Drawn from the "Fruit and vegetable disease - healthy vs rotten" set, we kept only mango images and discarded others.
- *Lemon Dataset:* Pulled from the "Lemon Quality Dataset," including pictures of good and poor-quality lemons, with folders renamed "lemon_good" and "lemon_bad" for clarity.

We combined these into a single dataset by sorting into folders by produce type and quality. By naming them as "mango_good," and "mango_bad,".

B. Preprocessing

Before moving forward, we prepped the images to keep them consistent and ready for analysis:

- *Resizing:* Adjusted every image to 128 × 128 pixels.
- *Gamma Correction*: Shifted brightness levels to handle uneven lighting and aid segmentation.
- *Color Space Conversion: Changed images to HSV format to better capture details.*

C. Segmentation and Background Removal

Removing the background accurately was vital for proper classification. We used these steps:

- *HSV Thresholding:* Converted images to HSV and set threshold ranges to separate produce from its surroundings.
- *Morphological Operations:* Applied Gaussian blur to smooth things out, then used opening, closing, and dilation to fine-tune the mask.
- *Edge Refinement:* Used Canny edge detection to define object edges clearly.
- *Convex Hull Masking:* Focused on the largest contour to ensure only the produce stayed in the frame.

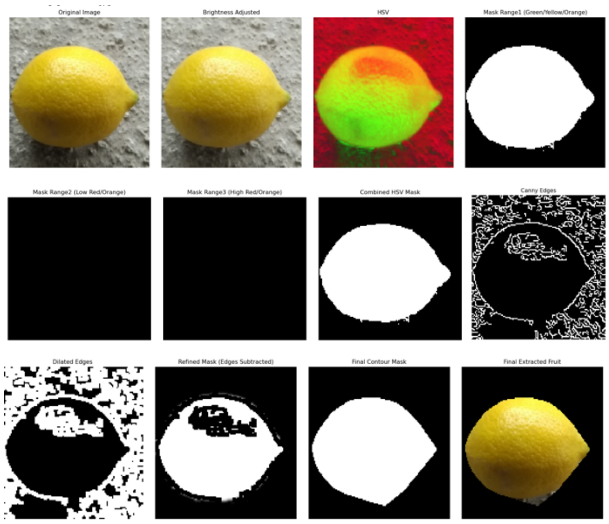

Fig. 2. Shows examples of segmented produce with and without background, demonstrating the effectiveness of HSV thresholding and morphological operations.

D. Feature Extraction

We pulled out specific traits to measure color, texture, and structure:

- *Color Features:* Created HSV histograms to track dominant colors.
- *Edge Density (Sobel Operator):* Used the Sobel operator to count edge pixels and assess surface roughness.
- *Shannon Entropy:* Calculated randomness in the image to judge texture complexity.

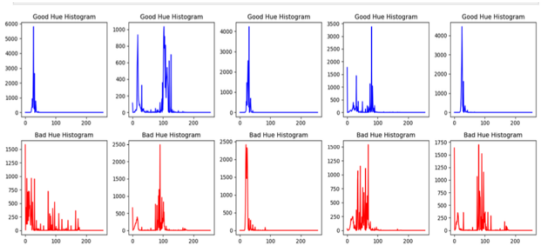

Fig. 3. Showcases the distribution of HSV histograms across good and bad produce.

E. Classification and Evaluation

A rule-based classifier was implemented based on extracted features:

- *Threshold-Based Decision:* If Edge Density > 0.1 or Entropy > 5.0, classify as Bad; Otherwise, classify as Good.
- *Dataset Split:* Training (60%), validation (20%), and testing (20%) sets.
- *Evaluation Metrics:* Accuracy scores, confusion matrices, and classification reports.

IV. Results and Evaluation

The experiments were conducted on the merged dataset of lemons and mangoes. The system achieved a classification accuracy of approximately 63–65% across training, validation, and test sets.

A. Performance Metrics

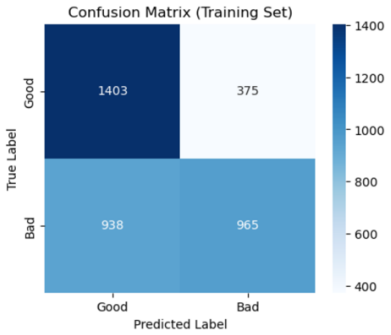

Fig. 4. Classification performance of the model on the training dataset.

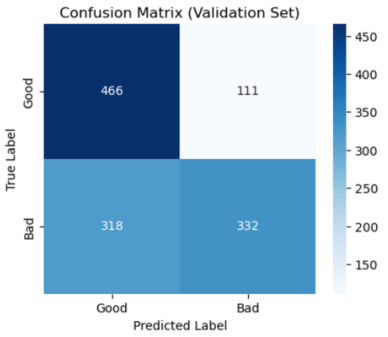

Fig. 5. Showcases the model's validation performance.

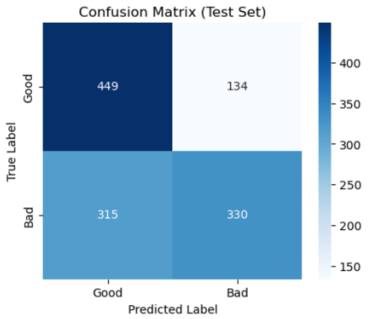

Fig. 6. llustrates how the model generalizes on unseen test data.

TABLE I. SUMMARIZES THE ACCURACY ACHIEVED ON TRAINING, VALIDATION, AND TEST SETS.

Dataset	Accuracy (%)
Training Set	64.33%
Validation Set	65.04%
Test Set	63.44%

The confusion matrices in Figs. 4, 5, and 6 illustrate model performance across training, validation, and test sets, respectively.

B. *Failure Cases*

Some items did not sort correctly, and we traced it back to these issues:

- *Lighting Variations:* Shadows or bright patches skewed the HSV thresholds.
- *Texture Misinterpretations:* Poor-quality produce with smooth surfaces got flagged as "Good" by mistake.
- *Background Interference:* Patchy segmentation let background bits confuse the process.
- *Color Similarity Confusion:* Yellowish bad lemons blended too closely with fresh ones.

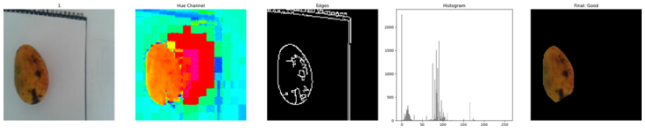

Fig. 7. Misclassification Example

C. *Success Cases*

Here are some correctly classified images as shown below in “Fig. 8”.

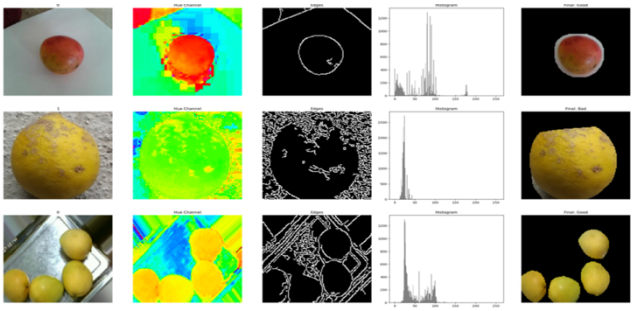

Fig. 8. Correctly classified Examples

V. CONCLUSION

This study crafted a cost-effective grading system for fruits and vegetables using classical image processing. It worked well for removing backgrounds, analyzing features, and sorting under steady conditions. Still, shifting light and texture mix-ups posed challenges. Moving forward, we could refine it by:

- *Adding Machine Learning Models:* Pairing our features with tools like support vector machines (SVM) or decision trees to boost accuracy.
- *Flexible Thresholding:* Adjusting segmentation dynamically for different light settings.
- *Better texture tools:* Using methods like Gabor filters or Local Binary Patterns (LBP) to sharpen texture analysis.

These results show that classical image processing offers a solid, affordable way to grade produce, though folding in advanced techniques could push it further.

REFERENCES

[1] R. Yusianto, R. Pradana, and A. Rachmat, “Smart potato grading using image processing and fuzzy grading system,” in Proc. Int. Seminar Appl. Technol. Inf. Commun. (iSemantic), Semarang, Indonesia, 2022, pp. 333–337. DOI: 10.1109/iSemantic55962.2022.9920382

[2] A. Baloch, A. Okatan, M. U. R. Jamali, N. Kanasro, M. Baloch, and A. Jamali, “The quality analysis of food and vegetable from image processing,” VAWKUM Trans. Comput. Sci., vol. 11, no. 2, pp. 1–17, 2023. DOI: 10.21015/vtcs.v11i2.1582.

[3] D. Jin, “Develop an olive-based grading algorithm using image processing,” Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 4, 2023. DOI: 10.14569/IJACSA.2023.0140483.

[4] L. D. Hanh and D. N. T. Bao, “Autonomous lemon grading system by using machine learning and traditional image processing,” Int. J. Interact. Des. Manuf., vol. 17, no. 1, pp. 445–452, 2023. DOI: 10.1007/s12008-022-00926-w.

[5] F. Guo and Q. Cao, “Study on color image processing-based intelligent fruit sorting system,” in Proc. 5th World Congr. Intell. Control Autom. (WCICA), Hangzhou, China, 2004, vol. 6, pp. 4802–4805. DOI: 10.1109/WCICA.2004.1343622.

[6] M. Subhan, “Fruit and vegetable disease healthy vs. rotten,” Kaggle, Accessed: Mar. 2025. [Online]. Available: https://www.kaggle.com/datasets/muhammad0subhan/fruit-and-vegetable-disease-healthy-vs-rotten.

[7] Y. Emir, “Lemon quality dataset,” Kaggle, Accessed: Mar. 2025. [Online].Available: https://www.kaggle.com/datasets/yusufemir/lemon-quality-dataset