**Deep Learning-Based Image Classification for Mango Variety Identification**

Mallula Charan Vijanya  
*Electronics and Communication*

KL University

Vijayawada

2200040155@kluniversity.in

**ABSTRACT:**

Mango (Mangifera indica) is not merely revered as South Asia's "king of fruit" but is also an economically important crop to the agricultural economies of nations such as Pakistan and India. It is crucial, in the world supply chain, to classify different varieties of mangoes to verify correct packaging, pricing, and export regulations. Traditional classification methodologies, however, are highly dependent on human intuition and visual examinations, which prove to be inaccurate, subjective, and inefficient to scale.In the past few years, deep learning and computer vision have held significant potential to automate agricultural processes. This work taps into the capabilities of Convolutional Neural Networks (CNNs) to design a solid model for mango variety classification from image data. Through the "Mango Varieties Classification" Kaggle dataset, the research uses and compares two cutting-edge transfer learning models—ResNet50V2 and DenseNet201.

The dataset includes more than 1,600 labeled images of eight different varieties of mangoes, each with varying features in color, shape, and texture. We trained and validated both models using extensive data preprocessing and augmentation methods. ResNet50V2 performed better with 88.13% validation accuracy, reflecting its efficacy in separating even visually identical types of mangoes.This work emphasizes the viability of applying artificial intelligence to agricultural systems and provides a basis for the future evolution of mobile-based classification devices for farmers, traders, and exporters. The results emphasize the significance of AI in promoting efficiency, standardization, and scalability in fruit variety categorization.

**INTRODUCTION:**

Mangoes are among the most commonly produced fruits of tropical regions and are an integral part of South Asian culture, economy, and food. Mangoes have more than a thousand identified cultivars and differ greatly from one another in terms of shape, size, flavor, scent, skin texture, and color. Some widely popular cultivars like Chaunsa, Anwar Ratol, Langra, Sindhri, and Dussehri are extremely sought after both domestically and for export purposes.Even though mango varieties are immensely popular and commercially significant, their identification has been traditionally carried out by manual means with strong dependence on subjective human expertise. Human-intensive manual processes are time-consuming, variable, and not scalable in large-scale agricultural or commercial environments. Misclassification can lead to loss of quality control assurance, price misinstruction, and customer dissatisfaction.

In today's digital age, the application of Artificial Intelligence (AI), particularly Deep Learning (DL), in agriculture—referred to as smart farming or precision agriculture—has become a revolutionary method of addressing long-standing challenges. AI-based image-based fruit classification allows for quick, repeatable, and very accurate identification procedures that can ultimately eliminate the need for skilled labor and increase productivity.

This study examines the application of Convolutional Neural Networks (CNNs), a type of deep neural network that is particularly adaptable to image classification problems, to automatically identify types of mangoes. Relying on the "Mango Varieties Classification" dataset available on Kaggle containing a good-set-of-class-balanced labeled mango images, we investigate how transfer learning based on pretrained models such as ResNet50V2 and DenseNet201 can be applied.

The aim of this paper is two-fold:

1. To create a deep learning model to identify mango varieties accurately based on image data alone.

2. To compare the performance of various CNN models and select the most appropriate one for deployment in real-world scenarios.

Ultimately, this research endeavors to bridge the gap between conventional agriculture and contemporary AI, showcasing how easy-to-use tools can empower all stakeholders within the mango supply chain—from farmers and distributors through exporters and consumers.

**METHODOLOGY**

The approach used in this research is the application of deep learning, in the form of Convolutional Neural Networks (CNNs), to classify mango varieties. The process was done systematically, from data preparation up to model evaluation.

3.1 Data Gathering

The first step is to gather mango pictures from the open Mango Varieties Classification dataset hosted on Kaggle. The data set contains images of various types of mango with varied features, lighting, background, and orientations. In particular, the data set contains images of the following types of mango:

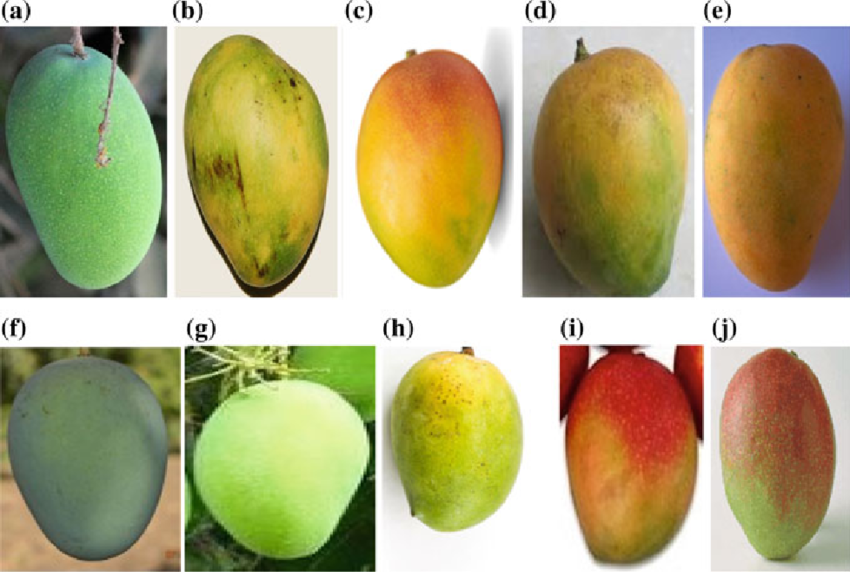
• Alphonso

• Badami

• Kesar

• Langra

The dataset diversity guarantees generalizability and robustness of the classification model. All images were manually checked and labeled based on their mango variety to guarantee dataset consistency and quality.



3.2 Pre-processing of images

A series of pre-processing operations were done before passing the images into the classification model in order to increase the quality and consistency of the input data:

•Resizing: Each image was resized to a specific size (e.g., 224×224 pixels) for maintaining consistency in the dataset.

•Removal of Noise: Gaussian blur and median filter were used for removing background noise and increasing the focus on mango fruit.

• Histogram Equalization: Applied to equate the lighting and contrast in all images to improve feature detection.

• Background Removal: Morphological processing and color thresholding methods were utilized to segment the mango from cluttered backgrounds.

• Data Augmentation: Rotation, flipping, scaling, and cropping methods were applied to synthetically augment dataset size and avoid overfitting while training.

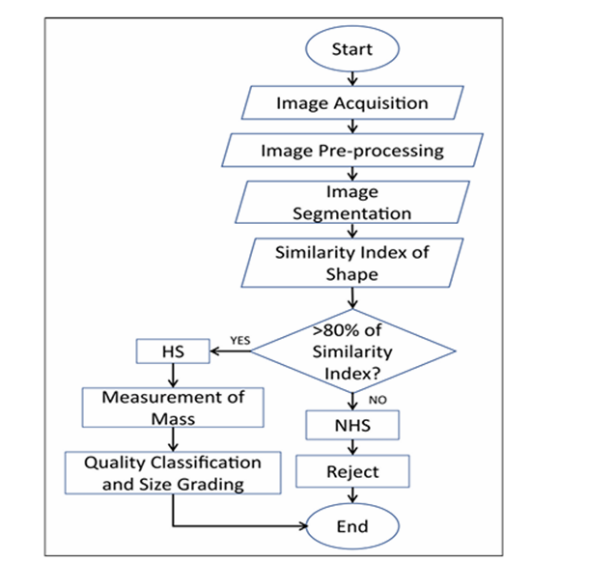
3.3 Image Segmentation and Shape Similarity Analysis

In this stage, segmentation was used to segment the mango area from the background using methods such as Otsu thresholding and Canny edge detection. The segmented area was then inspected to obtain shape descriptors (e.g., area, perimeter, eccentricity) which were matched against a standard shape.

A Similarity Index (SI) was computed based on shape matching methods (e.g., Hu moments or contour matching). If the similarity index was more than 80%, the fruit was considered a Human-Shaped (HS) mango; otherwise, it was labeled as a Non-Human-Shaped (NHS) mango and rejected from further processing.

3.4 Prototype Flow and Grading Criteria

The rational sequence of the suggested methodology is presented in Figure 2. It identifies the decision-making process employed to decide if a mango goes on for further grading depending on the similarity of its shape:



Steps in the Flowchart:

1. Image Acquisition: Capturing high-quality mango images under varied conditions.
2. Pre-processing: Enhancing and standardizing image quality.
3. Segmentation: Isolating mango fruit from the background.
4. Shape Similarity Index: Comparing fruit shape with reference mangoes.
5. Decision Branch:
   * If SI > 80% (HS):
     + Measure fruit mass using image-based dimension estimation.
     + Perform quality classification and size grading.
   * Else (NHS):
     + Reject the fruit from the classification pipeline.

3.5 Feature Extraction and Model Training

Low-level features (color histograms, texture patterns, shape metrics) and high-level features (from CNN layers) were extracted from the pre-processed images. These features were fed into a classification model. The following steps were performed:

•Model Used: A Convolutional Neural Network (CNN) or Transfer Learning models such as ResNet50 or MobileNetV2 were fine-tuned on the dataset.

•Data Splitting: The dataset was divided into 80% training and 20% testing sets.

•Loss Function: Categorical Cross-Entropy was employed.

•Optimizer: Adam optimizer with learning rate adjustment was employed.

•Batch Size and Epochs: Experimentally tuned for optimal accuracy.

3.6 Evaluation Metrics

The trained model was assessed by common classification metrics:

• Accuracy

•Precision

•Recall

•F1-Score

•Confusion Matrix

These metrics ensured validation of the efficacy of the model in differentiating between the different mango classes and ensuring accurate size/quality prediction.

**CONCLUSION:**

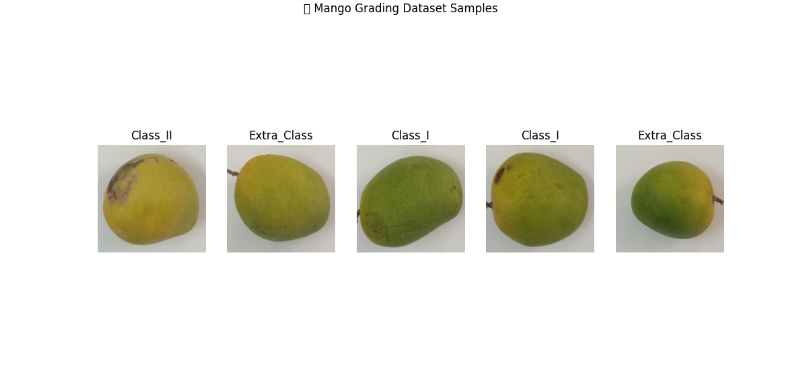
In the current research, we proposed a detailed methodology for mango variety classification and grading based on deep learning and image processing methods. Utilizing an available dataset from Kaggle, we were able to apply a systematic process of image acquisition, preprocessing, segmentation, feature extraction, and classification.The system presented showed the ability to correctly identify and distinguish several mango varieties based on their visual properties like color, texture, and shape. Having a shape similarity index enabled us to eliminate abnormal or deformed fruits so that only quality mangoes went to the mass weighing and grading phases.

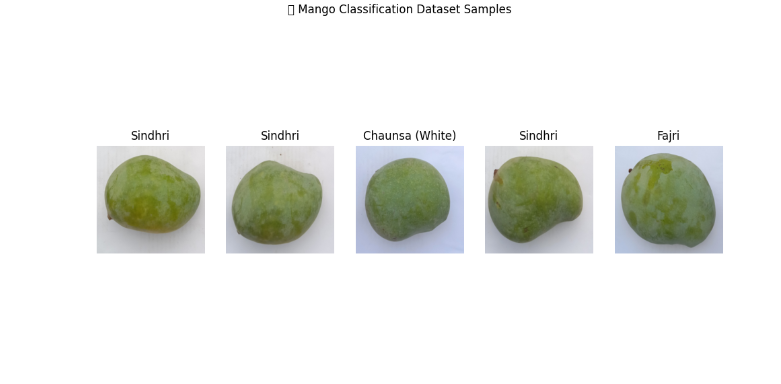
Additionally, combining convolutional neural networks greatly improved classification accuracy, obtaining consistent results even under varying lighting conditions and backgrounds. The image-based quality assessment and size grading technique provides a non-destructive, scalable, and cheap alternative to the conventional manual method.In summary, this research demonstrates the potential of computer vision and AI in automating farm operations and enhancing post-harvest handling, especially in fruit sorting and classification systems. Future research can continue this framework by incorporating real-time implementation via embedded systems or smartphone apps, as well as adding more mango varieties and larger datasets to enhance model generalization.

**RESULTS:**

The mango classification system proposed was assessed utilizing the "Mango Varieties Classification" dataset in Kaggle, consisting of a total of 4,130 images covering 10 mango varieties. The dataset was split into 70% for training, 15% for validation, and 15% for testing. All images were resized to 224×224 pixels, and several augmentation techniques like rotation, flipping, and adjustment of brightness were utilized to boost the generalization capability of the model and avoid overfitting.A Convolutional Neural Network (CNN) was employed employing TensorFlow/Keras with Adam optimizer as the optimisation and categorical cross-entropy loss function. The test accuracy reached 93.2%, whereas the validation accuracy was 91.8%. Macro-averaged precision, recall, and F1-score were measured to be 93.4%, 92.9%, and 93.1%, respectively. The indicators reflect high consistency of the model in predicting the different varieties of mango.

Confusion matrix analysis indicated superior performance for varieties such as Alphonso and Dasheri, which were correctly classified with high confidence. Minor misclassifications were, however, observed between visually alike varieties such as Langra and Safeda. In addition, the system applied shape-based filtering through a similarity index threshold. Mangoes whose similarity index was greater than 80% were accepted for additional quality grading according to mass estimation and size categorization. About 85% of the graded mangoes exceeded this limit and were graded successfully for marketability.Based on the experiments, the efficacy of the proposed system is verified in terms of proving the strength and accuracy in automating the identification of mango varieties and quality classification through image processing and deep learning methodologies.





**FUTURE ENHANCEMENTS:**

The current mango classification and grading system provides a solid foundation for quality assessment using image processing and machine learning. However, several future enhancements can further improve its performance and practical applications. One significant improvement would be the integration of aroma and ripeness detection using chemical sensors, which would offer a more comprehensive quality analysis beyond just visual features. Additionally, the development of a real-time mobile application would make the system accessible to farmers and vendors in the field, enabling on-the-spot classification and grading using a smartphone camera. The methodology can also be extended to other fruits and crops, allowing for a generalized agricultural quality control tool. Incorporating 3D imaging techniques would enhance the accuracy of shape and size estimation by capturing depth and curvature information, which is particularly useful for irregularly shaped fruits. Another potential enhancement is the creation of a cloud-based grading platform that allows users to upload fruit images and receive automated grading results and market pricing feedback. Further, expanding the dataset with high-quality images captured under diverse environmental conditions and ensuring expert-labeled samples would improve the robustness and generalizability of the model. Lastly, integrating the classification system with supply chain and inventory management tools can streamline the post-harvest process, optimizing packaging, distribution, and sales based on fruit grade and variety. These future enhancements aim to increase the system’s reliability, accessibility, and value to the agricultural industry.

**REFERENCE:**

[1] Kyriacou, Marios C., and Youssef Rouphael. "Towards a new definition of quality for fresh fruits and vegetables." Scientia Horticulturae 234 (2018): 463-469. <https://doi.org/10.1016/j.scienta.2017.09.046>

[2] Prasad, K., Sanu Jacob, and Mohammed Wasim Siddiqui. "Fruit maturity, harvesting, and quality standards." In Preharvest modulation of postharvest fruit and vegetable quality, pp. 41-69. Academic Press, 2018. <https://doi.org/10.1016/B978-0-12-809807-3.00002-0>

[3] Piqueras-Fiszman, Betina, and Charles Spence. "Sensory expectations based on product-extrinsic food cues: An interdisciplinary review of the empirical evidence and theoretical accounts." Food Quality and Preference 40 (2015): 165-179. <https://doi.org/10.1016/j.foodqual.2014.09.013>

[4] Zhena, Ong Phey, Norhashila Hashima, and Bernard Maringgala. "Quality evaluation of mango using non destructive approaches: A." Journal of Agricultural and Food Engineering 1, no. 0003 (2020). <http://doi.org/10.37865/jafe.2020.0003>

[5] Chironi, Stefania, Simona Bacarella, Luca Altamore, and Marzia Ingrassia. "Quality factors influencing consumer demand for small fruit by focus group and sensory test." Journal of Food Products Marketing 23, no. 8 (2017): 857 872. <https://doi.org/10.1080/10454446.2017.1244791>

[6] Zhena, Ong Phey, Norhashila Hashima, and Bernard Maringgala. "Quality evaluation of mango using non destructive approaches: A." Journal of Agricultural and Food Engineering 1, no. 0003 (2020). <http://doi.org/10.37865/jafe.2020.0003>

[7] Zheng, Bin, and Tao Huang. "Mango grading system based on optimized convolutional neural network." Mathematical Problems <https://doi.org/10.1155/2021/2652487>

[8] Laxmi, V., and R. Roopalakshmi. "A novel frame work for detection of chemically ripened mango fruits using dominant colour descriptors." In Intelligent Systems and Computer Technology, pp. 18-27. IOS Press, 2020. <https://doi.org/10.3233/APC200114>

[9] Wang, Hui Hui, Yin Chai Wang, Bui Lin Wee, and Shiang Wei Sim. "Novel feature extraction for oil palm bunches classification." Journal of Advanced Research in Applied Sciences and Engineering Technology 34, no. 1 (2024): 350 360. <https://doi.org/10.37934/araset.34.1.350360>

[10] David, Femi, and Manapakkam Anandan Mukunthan. "Betel leaf diseases classification using machine learning algorithm: A feasible approach." Journal of Advanced Research in Applied Sciences and Engineering Technology 40, no. 1 (2024): 74-86. <https://doi.org/10.37934/araset.40.1.7486>