TSSA Journal of Tomography System & Sensors Application Vol. 7, Issue 1, 2024
www.tssa.com.my e-ISSN: 636-9133

# Chokanan Mango Fruit Maturity Detection using K-Nearest Neighbor

Nur Athirah Syafiqah Noramli[1*], Hajar Izzati Mohd Ghazalli[1], Herlina Abdul Rahim[2],

[1*] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 77300 Merlimau Melaka, Malaysia
[2] Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor. Malaysia

Corresponding author* athirah.noramli1@gmail.com
Available online 30 June 2024

**ABSTRACT**

This research introduces a technique utilizing machine learning, specifically the K-Nearest Neighbors (KNN) algorithm in Python, to determine the maturity of mango fruit. The main goal is to create a system capable of precisely evaluating mango fruit maturity, which is essential for improving post-harvest processes and ensuring top-notch produce for consumers. The proposed method involves extracting pertinent features from mango fruit images and training the KNN classifier with labeled data. Subsequently, the trained model is deployed to categorize unseen mango samples based on their maturity stages. Experimental findings validate the efficacy of the developed system in accurately assessing mango fruit maturity, achieving high classification accuracy. This study significantly contributes to fruit quality assessment methodologies, offering a practical solution for the fruit industry to enhance operational efficiency and meet consumer expectations.

**Keywords:** Mango fruit, k-nearest Neighbors, maturity

## 1. Introduction

In Malaysian culture, mangoes hold a significant place, with their cultivation and consumption rooted in the influence of Indian traders and immigrants who introduced various mango varieties to the region. Today, Malaysia boasts a diverse range of mango types like Harumanis, Chokanan, and Tommy Atkins, each offering distinct flavors and characteristics. Identifying the ideal ripeness of mangoes is crucial for savoring their full sweetness and aroma, typically signaled by their vibrant yellow-orange skin tinged with red and a tender, juicy flesh [1].

Mangoes, being tropical fruits, undergo notable changes as they mature. Their ripeness is discerned by factors such as the color, texture, and fragrance of the skin [2]. A ripe mango typically exhibits a bright yellow-orange skin with a subtle reddish hue. Its flesh is soft, succulent, and emits a delightful aroma. It's important to allow mangoes to ripen fully before consumption or culinary use to ensure optimal sweetness and flavor.

Current research on using machine learning to recognize mango fruit maturity serves multiple purposes, addressing key challenges encountered by the mango industry. Firstly, it aims to tackle the issue of determining the optimal harvest time. Harvesting mangoes when they reach the ideal stage of maturity is essential for ensuring desirable taste, texture, aroma, and overall quality. This comprehensive approach facilitates a thorough assessment of mango maturity, enabling farmers and producers to harvest the fruit at its peak, thereby maximizing its flavor and nutritional content [3]. Additionally, the research endeavors to enhance post-harvest handling techniques. After harvesting, mangoes continue to ripen, necessitating careful management to maintain optimal shelf life and quality. By leveraging machine learning methods, researchers can develop models that forecast the ripening process of mangoes based on their maturity level at harvest, environmental conditions, and storage conditions. This knowledge empowers the industry to mitigate post-harvest losses, optimize supply chain operations, and ultimately deliver top-quality mangoes to consumers [4].

## 2. Background

A various study that have utilized computer vision and machine learning techniques to assess fruit ripeness. Kangune et al. conducted research on grape maturity classification employing CNN and SVM. Their dataset consisted of 4000 images, evenly split between unripen and ripened grapes, with preprocessing involving color and morphological features, including Gaussian blur. The CNN model achieved an accuracy of 79.49%, surpassing SVM's accuracy of 69% [5]. In another study in 2019, Mazen & Nashat aimed to classify banana maturity using multiple machine learning techniques. Their dataset comprised different stages of banana ripeness, and they achieved high accuracy rates, notably 100% for green and overripe bananas and 97.75% for mid-ripe yellowish-green bananas [6]. Suban et al. utilized the KNN algorithm to classify Papaya Carica fruit ripeness, achieving 100% accuracy with a dataset of 12 images [7]. Momeny et

al. focused on cherry classification based on shape, employing CNN with various feature extraction techniques, achieving up to 99% accuracy [8]. In a study by Hamza & Chtourou, an ANN was developed to assess apple ripeness based on color, achieving a high precision rate of 96.66%.

The primary objective of this research is to achieve the highest classification accuracy for mango ripeness using the KNN algorithm. This decision is based on the KNN algorithm's superior performance, achieving 100% accuracy compared to other algorithms studied. Image preprocessing techniques such as noise elimination and scaling were employed, and data augmentation was applied to generate variants of the photos within the dataset.

## 3. Methodology

Figure 1 depicts the sequential stages of the study, encompassing literature review, dataset acquisition, image preprocessing, KNN model implementation, and evaluation. As per the illustration, the initial phase entails a comprehensive examination of existing literature. During this step, pertinent research documents and reports are collected from diverse academic journals, serving as foundational resources for the study. Through the literature review process, various approaches for classifying fruit ripeness are scrutinized and compared, facilitating the selection of the methodology employed in this study. KNN was specifically selected based on prior research highlighting its capability to attain high accuracy levels.

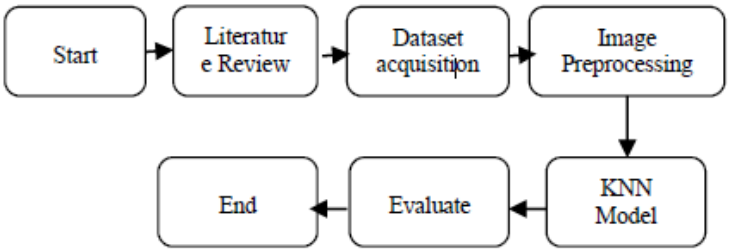


**Figure 1.** Research Methodology

The second phase involves acquiring the dataset, which comprised 500 images of Chokanan mango fruit obtained from Kaggle. These images were categorized into two ripeness classes: ripe and unripe, each containing 250 images. The image sizes varied from 225 x 225 pixels to 960 x 540 pixels. Following dataset acquisition, the next stage focused on image processing. Despite careful image capture, various factors could influence outcomes. To address this, preprocessing steps such as image enhancement and resizing were applied. Image enhancement adjusted pixel intensity to improve contrast, while resizing reduced image dimensions while preserving quality for efficient processing. During this stage, images exclusively contained relevant foreground, simplifying segmentation. Segmentation was achieved using the global thresholding method, determining a single threshold value to uniformly separate foreground (mango portion) from background.

Feature extraction played a pivotal role in image classification. In the context of mango maturity classification, significant features were extracted from segmented images to capture relevant information. In the referenced research, color features were extracted from each sample image, encompassing statistics such as mean, median, mode, standard deviation, and skewness for the red (R), green (G), and blue (B) color channels. These extracted features underwent further analysis, with the most important ones selected to enhance accuracy and precision in the classification process.

The KNN model was implemented at this stage, with 400 images (80% of the dataset) allocated for training, evenly split between ripe and unripe categories, each containing 200 images. Testing involved 100 images (20% of the dataset), with 50 images allocated to each ripe and unripe category. Despite its compact size, the KNN algorithm demonstrated remarkable image recognition accuracy, leading to its selection for developing a compact yet high-performing model for mango ripeness classification.

Following the implementation of the KNN model, the subsequent step involved evaluating its performance. This evaluation included assessing each model's performance by varying the value of k. Metrics such as accuracy, precision, recall, specificity, sensitivity, and F-measure were utilized to measure performance, ultimately determining the superior model.

### 3.1 Identification

In this study, the dataset comprised 500 images of Chokanan mango fruit sourced from Kaggle. These images were categorized into two ripeness classes: ripe and unripe, each containing 250 images. The image dimensions ranged from 225 x 225 pixels to 960 x 540 pixels. Examples of ripe and unripe mangoes are depicted in Figure 2 and Figure 3, respectively.

**Figure 2.** Unripe mango

**Figure 3.** Ripe mango

### 3.2 K-Nearest Neighbor Method

KNN is a simple algorithm that stores all available data and classifies new cases by comparing them to existing cases, typically using distance functions. Classification is based on a majority vote from neighboring cases. The choice of the value k determines the number of nearest neighbors considered, where k = 3 means sorting the three closest values. By employing Euclidean Distance, distances between training data points are calculated. To expedite the process, distances are arranged, enabling the classification of fruits as ripe or unripe based on the shortest distance.

### 3.3 Data Analysis

The mango maturity classification process consists of five stages, as depicted in Figure 4. It commences with retrieving the image from the database, followed by employing advanced image processing techniques to extract pertinent features like color, texture, and shape. These extracted features are subsequently inputted into a machine learning algorithm, which has been trained on a substantial dataset of labeled mango images, enabling accurate classification of the maturity stage.

Upon image retrieval from the database for the mango maturity classification system, the subsequent step is to determine the optimal value of K in the KNN algorithm. K denotes the number of nearest neighbors considered during predictions. This stage involves selecting an appropriate K value by evaluating the KNN algorithm's performance on a training dataset.

Techniques such as cross-validation are employed to ascertain the K value that produces the highest accuracy or minimizes error for mango maturity classification. The chosen K value is then utilized in the subsequent steps of the system to classify the maturity level of mangoes based on their retrieved images.

Following the determination of the optimal K value for mango maturity classification in the KNN algorithm, the next step entails computing Euclidean distances between the retrieved image and other dataset instances. Euclidean distance quantifies the similarity between features. This computation is conducted for the K nearest neighbors. Identifying the optimal K involves selecting the number of neighbors that yield the best performance, aiming to maximize accuracy. Once determined, this K value is employed in subsequent steps to classify mango maturity levels through majority voting based on nearest neighbors.

After applying the majority voting technique in the KNN algorithm for mango maturity classification, the subsequent step involves obtaining the final result with accuracy. Mangoes are designated as ripe or unripe based on the most frequent class among the K nearest neighbors. Once the majority class is determined, it is assigned as the predicted class for the mango. Accuracy is assessed by comparing predicted labels with actual ones in the dataset. This system proficiently classifies mangoes, assisting in decisions regarding harvesting, storage, or distribution processes.

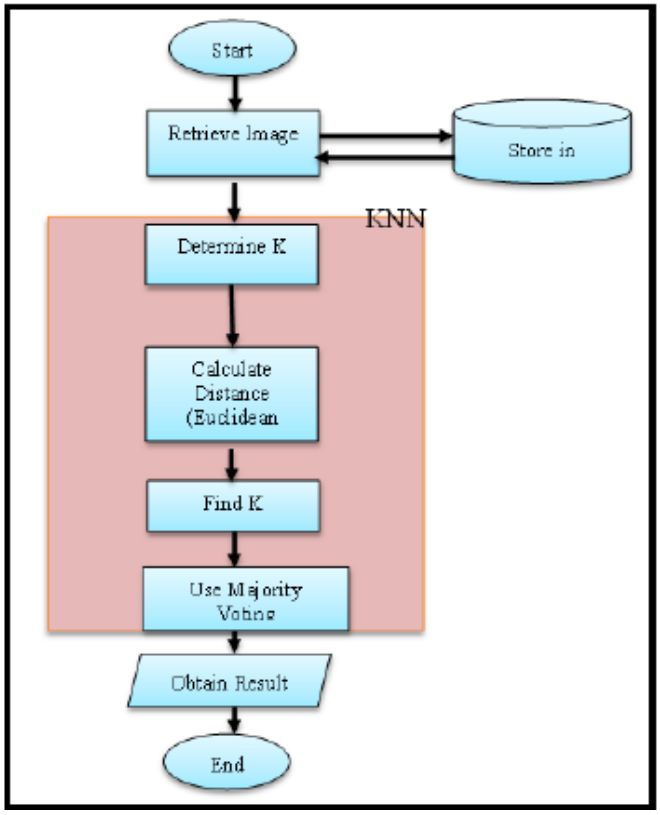


**Figure 4.** Flowchart of classification system

## 4. Results and Discussion

### 4.1 Data Collection

The study utilized a dataset comprising 500 images of Chokanan mango fruit, sourced from Kaggle as depicted in Figures 5 and 6. These images were categorized into two ripeness classes: ripe and unripe, with each class containing 250 images. The image dimensions ranged from 225 x 225 pixels to 960 x 540 pixels. The training set comprised 400 images (80% of the dataset), evenly split between ripe and unripe categories, with each containing 200 images. The testing set included 100 images (20% of the dataset), with 50 images allocated to each ripeness class.

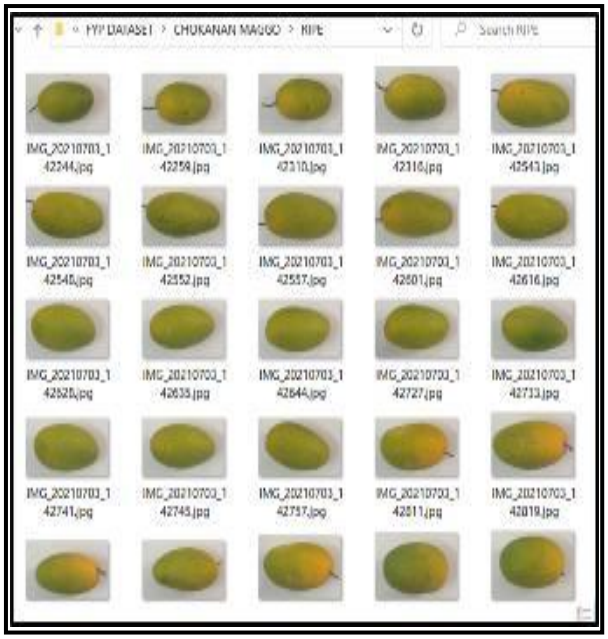

**Figure 5.** Ripe mangoes

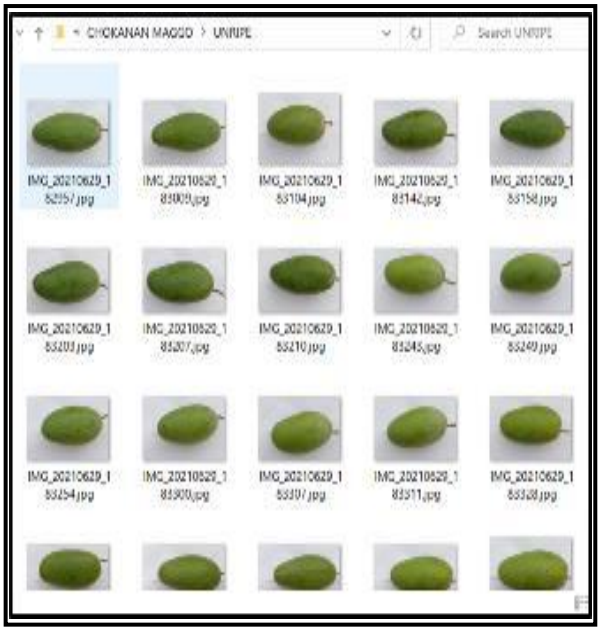

**Figure 6.** Unripe mangoes

### 4.2 Pre-Processing

Preprocessing encompassed both image enhancement and resizing. Enhancement techniques were employed to refine contrast by fine-tuning pixel intensity. Resizing operations were conducted to reduce image dimensions without compromising quality, aiming for improved processing efficiency. This preprocessing phase ensured that images primarily featured pertinent foreground elements, facilitating subsequent segmentation tasks. Segmentation utilized the global thresholding approach, which established a singular threshold value to uniformly differentiate foreground from background elements.

### 4.3 Classification using the K-Nearest Neighbor Method

The mango fruit maturity classification task employs a KNN classifier model. Data is fetched from a designated directory, and features are derived from images using color histogram and edge detection methods. Subsequently, the dataset is partitioned into training and testing subsets, allocating 80% for training and 20% for testing purposes. The KNN classifier is configured with five neighbors, utilizing the Euclidean distance metric and distance-based weighting, and then trained on the training dataset. The system demonstrates precise classification of unripe mango maturity, achieving a test accuracy of 99.01%, as depicted in Figure 7.

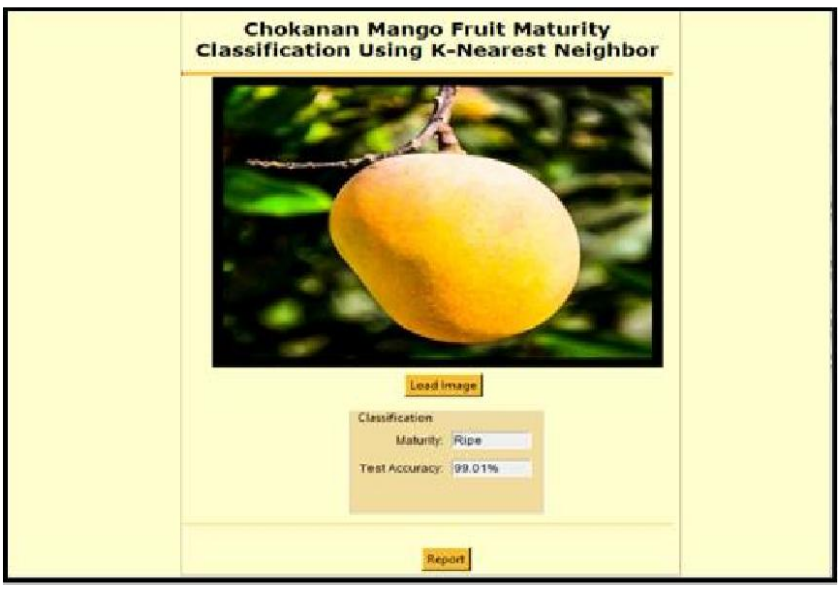


**Figure 7.** Classification

### 4.4 Comparison Dataset with Loaded Image

Figure 8 showcases color histograms extracted from the loaded image, enabling a direct comparison with histograms from various other images, including those within the loaded image (Figure 9), as well as from the training and testing datasets categorized as ripe or unripe (Figure 10). These histograms are visually distinguished by color: blue represents ripe, red denotes unripe, and green signifies the loaded image. Figure 8 further elucidates this feature, facilitating effortless differentiation and comparison of color distributions, thus providing a comprehensive visual depiction of color characteristics across diverse datasets. Importantly, when the loaded image predominantly exhibits characteristics of unripeness, it indirectly leads to its classification as unripe. This aspect bears significance in discerning the maturity level of Chokanan mangoes, offering valuable insights into the classification process.

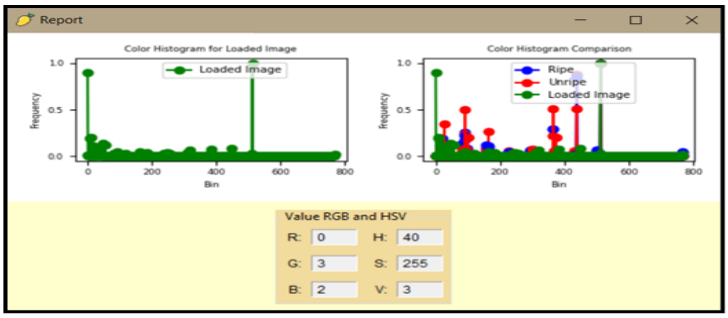


**Figure 8.** Histogram comparison

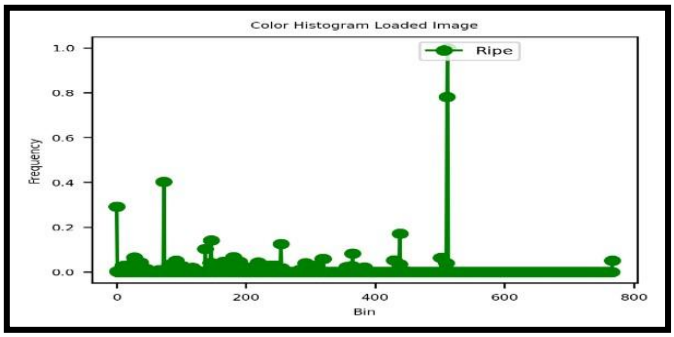


**Figure 9.** Loaded Image Histogram

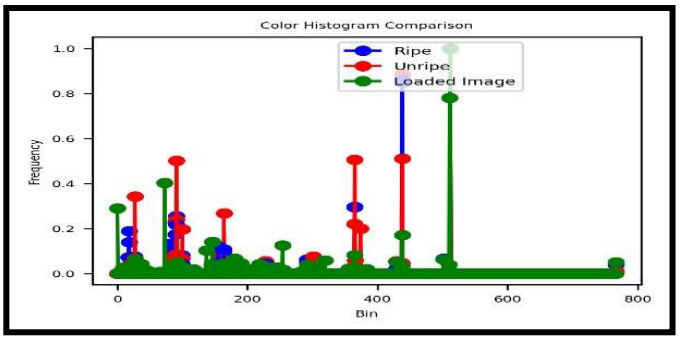


**Figure 10.** Testing, training and loaded image histogram

### 4.5 Accuracy of K-Nearest Neighbor Method

In this mango maturity classification project, the performance of the KNN classifier model was evaluated through both training and testing phases to gauge its accuracy. Simple Python code executed within the Sublime environment facilitated accuracy testing. Figure 11 presents the results, showcasing accuracy scores for both training and testing, along with confusion matrices for each. The accuracy testing phase yielded an impressive score of 100.00%, indicating the model's exceptional ability to classify mangoes into "Ripe" and "Unripe" categories. Moreover, the Classification Report provided metrics such as precision, recall, and f1-score for both the Testing and Training Sets, further validating the model's efficacy. Overall, these findings underscore the model's robust performance, characterized by consistently high accuracy rates across both training and testing datasets.

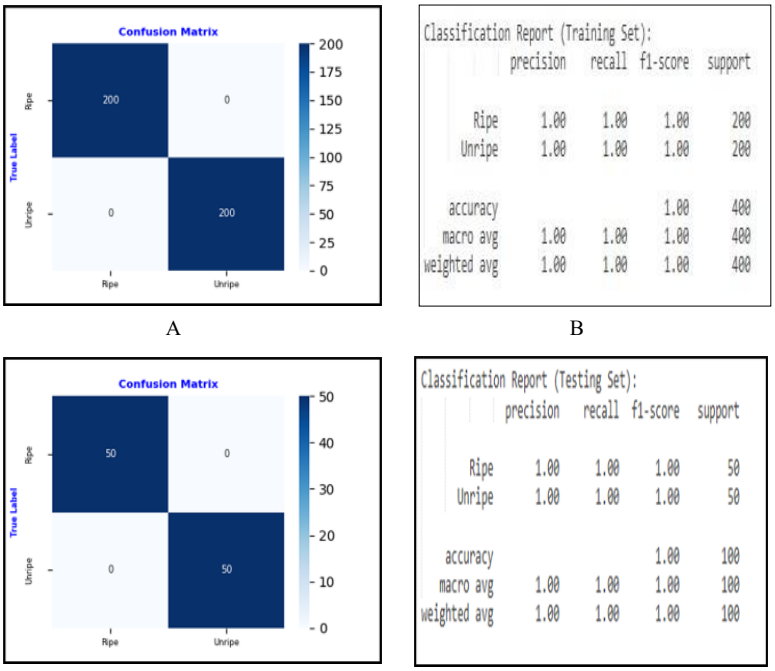


**Figure 11.** Confusion Matrix and Classification Report with different Model Performance, (A) Confusion Matrix of Training, (B) Classification Report of Training set, (C) Confusion Matrix of Testing, (D) Classification Report of Testing set

### 4.6 Comparison with Different k Values

Table 1 provides a comparative analysis of the KNN algorithm's performance across diverse hyperparameter configurations, encompassing variations in k values, distance metrics (Euclidean or Manhattan), weighting schemes (Uniform or Distance), and the size of the mango dataset (500 samples). Generally, smaller k values, such as 3, tend to yield higher accuracy rates on both testing and training sets, whereas larger k values may result in diminished accuracy. The selection of distance metric significantly influences performance, with instances where either Euclidean or Manhattan distance outperforms the other depending on the context. Weighting schemes also exert an impact on accuracy, with distance-based weighting typically leading to higher accuracy, particularly when combined with smaller k values. Among the configurations assessed, the setup featuring k=5, Euclidean distance, and distance-based weighting emerges as the most noteworthy, exhibiting flawless accuracy across both datasets and thereby representing the optimal choice for mango maturity classification.

Table 1 Hyperparameter Comparison

| Value k | Distance Metrics | Weighting | Accuracy (Testing) | Accuracy (Training) |
|---|---|---|---|---|
| 3 | Euclidean | Uniform | 90.5% | 97.5% |
| 3 | - | - | 94.5% | 99.2% |
| 4 | Euclidean | Distance | 73.5% | 80.3% |
| 4 | Manhattan | Uniform | 88.5% | 90.7% |
| 4 | Euclidean | Uniform | 87.5% | 90.7% |
| 5 | - | Uniform | 100.0% | 100.0% |
| 5 | - | Distance | 95.0% | 97.0% |
| 5 | Manhattan | Uniform | 89.5% | 95.8% |
| 7 | Euclidean | Distance | 83.5% | 92.2% |
| 7 | - | - | 86.5% | 94.7% |
| 7 | Euclidean | Uniform | 93.5% | 98.8% |
| 7 | Manhattan | Distance | 89.5% | 95.4% |
| 9 | - | - | 76.5% | 89.6% |
| 9 | Manhattan | Uniform | 77.5% | 89.7% |

### 4.7 Sample Input and Output Dataset

To ensure accurate classification of mango fruit features, a KNN architecture was utilized to analyze 10 images for each of the 2 categories, reflecting distinct stages of mango maturity. Below, Table 2 displays the results obtained from categorizing the input images.

Table 2 Sample Input and Output for Test and Train Dataset

| No | Input | Actual Input | Match Output | Accuracy | Pass /Fail |
|---|---|---|---|---|---|
| 1 |  | Ripe | Ripe | 100% | Pass |
| 2 |  | Ripe | Ripe | 100% | Pass |
| 3 |  | Ripe | Ripe | 100% | Pass |
| 4 |  | Ripe | Ripe | 100% | Pass |
| 5 |  | Ripe | Ripe | 100% | Pass |
| 16 |  | Unripe | Unripe | 100% | Pass |
| 17 |  | Unripe | Unripe | 100% | Pass |
| 18 |  | Unripe | Unripe | 100% | Pass |
| 19 |  | Unripe | Unripe | 100% | Pass |
| 20 |  | Unripe | Unripe | 100% | Pass |

## 5. Conclusion

In conclusion, this study has effectively fulfilled its primary aim of developing a comprehensive system capable of discerning the ripeness of Chokanan mangoes. Through meticulous planning and detailed documentation, the project seamlessly translated requirements into specific system functionalities. The secondary objective, to devise a machine learning-based system for precisely categorizing mango ripeness, was successfully realized with the implementation of a user-friendly web interface. The assessment of the system's accuracy using the KNN classification technique revealed an impressive 100% accuracy rate, underscoring its reliability and effectiveness. In summary, this system holds significant promise for the food and agriculture sectors, as it combines cutting-edge technology, thoughtful design, and rigorous testing to deliver practical benefits.

## 6. References

[1] T. Lawson, G. W. Lycett, A. Ali, and F. Chin, "Characterization of Southeast Asia mangoes (Mangifera indica L) according to their physicochemical attributes," 2018, doi: 10.1016/j.scienta.2018.08.014.

[2] K. A. B. Ahmad, M. Othman, S. L. Syed Abdullah, N. Rasidah Ali, and S. R. Muhamat Dawam, "Mango Shape Maturity Classification Using Image Processing," ICRAIE 2019 - 4th International Conference and Workshops on Recent Advances and Innovations in Engineering: Thriving Technologies, Nov. 2019, doi: 10.1109/ICRAIE47735.2019.9037776.

[3] S. M. T. Islam, M. Nurullah, and M. Samsuzzaman, "Mango Fruit's Maturity Status Specification Based on Machine Learning using Image Processing," 2020 IEEE Region 10 Symposium, TENSYMP 2020, pp. 1355–1358, Jun. 2020, doi: 10.1109/TENSYMP50017.2020.9230951.

[4] M. Khojastehnazhand, V. Mohammadi, and S. Minaei, "Maturity detection and volume estimation of apricot using image processing technique," Sci Hortic, vol. 251, pp. 247–251, Jun. 2019, doi: 10.1016/J.SCIENTA.2019.03.033.

[5] K. Kangune, V. Kulkarni, and P. Kosamkar, "Grapes Ripeness Estimation using Convolutional Neural network and Support Vector Machine," 2019 Global Conference for Advancement in Technology, GCAT 2019, Oct. 2019, doi: 10.1109/GCAT47503.2019.8978341.

[6] F. M. A. Mazen and A. A. Nashat, "Ripeness Classification of Bananas Using an Artificial Neural Network," Arab J Sci Eng, vol. 44, no. 8, pp. 6901–6910, Aug. 2019, doi: 10.1007/s13369-018-03695-5.

[7] I. B. Suban, A. Paramartha, M. Fortwonatus, and A. J. Santoso, "Identification the Maturity Level of Carica Papaya Using the K-Nearest Neighbor," in Journal of Physics: Conference Series, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1742-6596/1577/1/012028.

[8] M. Momeny, A. Jahanbakhshi, K. Jafarnezhad, and Y.-D. Zhang, "Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach," 2020, doi: 10.1016/j.postharvbio.2020.111204.