ARTIFICIAL INTELLIGENCE MODELS FOR VARIETY AND MATURITY CLASSIFICATION OF THAI COMMERCIAL MANGOES

PHANUPHONG SUTHAWAS

MASTER OF SCIENCE

IN

DIGITAL TRANSFORMATION TECHNOLOGY

SCHOOL OF APPLIED DIGITAL TECHNOLOGY

MAE FAH LUANG UNIVERSITY

2024

ARTIFICIAL INTELLIGENCE MODELS FOR VARIETY AND MATURITY CLASSIFICATION OF THAI COMMERCIAL MANGOES

PHANUPHONG SUTHAWAS

THIS THESIS IS A PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
DIGITAL TRANSFORMATION TECHNOLOGY

SCHOOL OF APPLIED DIGITAL TECHNOLOGY
MAE FAH LUANG UNIVERSITY
2024

THESIS APPROVAL

MAE FAH LUANG UNIVERSITY

FOR

MASTER OF SCIENCE

IN DIGITAL TRANSFORMATION TECHNOLOGY

Thesis Title: Artificial Intelligence Models for Variety and Maturity Classification of Thai Commercial Mangoes

Author: Phanuphong Suthawas

Examination Committee:

Assistant Professor Teerawat Kamnardsiri, Ph. D.	Chairperson
Assistant Professor Sujitra Arwatchananukul, Ph. D.	Member
Associate Professor Nattapol Aunsri, Ph. D.	Member
Assistant Professor Worasak Rueangsirarak, Ph. D.	Member
Assistant Professor Surapong Uttama, Ph. D.	Member

Advisors:

Sujitra A.

....Advisor

(Assistant Professor Sujitra Arwatchananukul, Ph. D.)

N. Au

....Co-Advisor

(Associate Professor Nattapol Aunsri, Ph. D.)

Dean:

....

(Assistant Professor Nacha Chondamrongkul, Ph. D.)

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to everyone who supported me throughout this thesis.

First and foremost, I sincerely thank my advisor, Asst. Prof. Sujitra Arwatchananukul, Ph. D., and my co-advisor, Assoc. Prof. Nattapol Aunsri, Ph. D., for their invaluable guidance, support, and the opportunity to work under their mentorship. Their patience, expertise, and encouragement have been crucial in shaping this work.

I also want to thank my colleagues for their camaraderie and the stimulating discussions we shared. Their support and collaboration have been a great source of inspiration.

Finally, I am deeply grateful to my family for their patience, understanding, and constant encouragement, which kept me motivated throughout this journey.

Phanuphong Suthawas

Thesis Title	Artificial Intelligence Models for Variety and Maturity Classification of Thai Commercial Mangoes
Author	Phanuphong Suthawas
Degree	Master of Science (Digital Transformation Technology)
Advisor	Assistant Professor Sujitra Arwatchananukul, Ph. D.
Co-Advisor	Associate Professor Nattapol Aunsri, Ph. D.

ABSTRACT

Thailand is one of the world's largest mango producers and exporters, where traditional grading methods rely on farmers assessing characteristics like color, texture, size, and shape. These methods, however, can be inconsistent. This study presents an AI-driven approach for automated mango classification, consisting of two stages: variety classification using a Random Forest classifier and maturity classification using machine learning and deep learning models. The Random Forest classifier, after hyperparameter tuning, achieved a remarkable accuracy of 99.63% for mango variety classification. Following this, mangoes are categorized into three maturity grades: Immaturity (M1), Exporting Maturity (M2), and Domestic Maturity (M3). The highest maturity classification accuracies were 80.00% for Mahachanok using InceptionV3, 84.40% for Namdokmai Sithong using Gradient Boosting, and 83.33% for R2E2 using Random Forest. Both models were integrated into a real-time web application, providing an efficient and scalable solution for mango classification, improving consistency and productivity in the agricultural sector.

Keywords: Maturity, Variety, Prediction, Machine Learning, Deep Learning, InceptionV3, Random Forest, Classification

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

ABBREVIATIONS AND SYMBOLS

MHN	Mahachanok
NDM	NamDokmai Sithong
R2E2	R2E2
M1	Immaturity
M2	Exporting Maturity
M3	Domestic Maturity
RF	Random Forest
GB	Gradient Boosting
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
DT	Decision Tree

CHAPTER 1

INTRODUCTION

1.1 Background and Problems

Mangoes, often referred to as the "King of Fruits," are of paramount importance to Thailand's agricultural sector (1,2). The mango (Mangifera indica L.), a member of the Anacardiaceae family, stands as one of the most significant tropical fruits globally, valued for its distinctive aroma, rich flavor, and exceptional nutritional content (3). As one of the world's leading producers and exporters of mangoes, Thailand plays a crucial role in the global mango market. In April 2024, Thailand exported approximately 13.17 million kilograms of fresh and frozen mangoes, with an export value of approximately 811.8 million Thai baht, reflecting a slight decline compared to previous months (4). The primary mango-producing regions in Thailand are located in the northern and northeastern provinces, particularly Chiang Mai, Lamphun, Sukhothai, and Loei. Among the most cultivated varieties in the country are Nam Dokmai Sithong (NDM), Chokanan, Khiou Sawoey, Khiou Moragot, R2E2, and Mahachanok (MHN) (5). Mangoes are highly perishable, with postharvest losses in developing countries ranging from 20% to 60% (6). These losses are influenced by factors such as fruit cultivar, storage conditions, water content, specific gravity, and maturity at harvest (7). The flavor, texture, and visual appeal of mangoes are significantly impacted by their maturity at harvest, which, in turn, affects their marketability. The key Thai mango varieties display distinct physical characteristics: NDM is oval-shaped and medium to large in size, MHN is cylindrical and medium-sized, while R2E2 is round and large. These varieties are highly prized in both domestic and international markets, making them prime candidates for advanced research and classification. To enhance mango classification and mitigate postharvest losses, this study aims to develop an automated mango classification system utilizing machine learning and advanced image analysis techniques. By integrating shape, texture, and color features, the proposed model will classify mangoes based on both variety and maturity stage. Additionally, a user-friendly web application will be developed to facilitate real-time classification, providing valuable support to farmers, distributors, and retailers in optimizing mango quality for both local and export markets.

1.2 Research Objectives

To develop an AI model for classifying Thai mango varieties Namdokmai Sithong, Mahachanok, and R2E2 and their maturity into three grades (M1: Immaturity, M2: Exporting Maturity, M3: Domestic Maturity), integrating it into a real-time web application.

1.3 Research Objectives

This research focuses on developing a comprehensive database for three prominent mango varieties NDM, MHN, and R2E2—utilizing machine learning techniques to classify mangoes into three maturity stages: Immaturity (M1), Exporting Maturity (M2), and Domestic Maturity (M3). M1 mangoes are in the early stages of ripening, lacking the necessary color, texture, and firmness for consumption or export, and remain underdeveloped in sweetness and flavor. M2 mangoes have reached optimal ripeness for international shipment, exhibiting a well-balanced combination of color, texture, and flavor while meeting strict export standards to ensure their quality during long-distance transportation. M3 mangoes are fully ripe for local consumption but may not meet the more rigorous export criteria, though they remain suitable for immediate sale in domestic markets. To enhance practical accessibility, the classification model is integrated into a user-friendly web application designed for farmers, distributors, and retailers, enabling real-time mango classification to ensure premium-quality mangoes are selected for export while optimizing the distribution of those designated for local markets.

1.4 Thesis Structure

The thesis structure is introduced in this section to outline the organization and sequence of the work. Each chapter is briefly summarized, highlighting the main content and objectives. This summary allows readers to gain a comprehensive overview of
the thesis and grasp the key components and progression of the research.

Chapter 2: Literature Review, this chapter presents a comprehensive review of relevant literature, including previous studies and research findings, to establish the theoretical foundation and contextualize the research.

Chapter 3: Research Methodology, this chapter presents the details of the research methodology including experimental procedures, data collection, data preprocessing, data modeling, and model evaluation to achieve the experiment result.

Chapter 4: Results, this chapter presents the results of the research study, which included accuracy, precision, recall, F1-score. These measures were used to assess the performance and predictive capabilities of the implemented model.

Chapter 5: Conclusion, this chapter synthesizes the outcomes of this study, offering information into their implications and providing recommendations for future research directions.

CHAPTER 2

LITERATURE REVIEWS

2.1 Mango Characteristics and Classification

This section provides an overview of current research on developing classification models for mango variety and maturity stage. With ongoing advancements in technology, the use of machine learning classifiers is steadily increasing to analyze and predict mango variety and maturity stages. This section aims to provide a comprehensive understanding of the current state of research in these areas.

2.1.1 Variety

The diversity of mango varieties contributes to variations in fruit characteristics such as size, shape, color, texture, sweetness, and fiber content. These attributes influence consumer preferences and determine the suitability of the fruit for different market segments, including fresh consumption, processing, and export. The following varieties are among the most commercially significant:

2.1.1.1 Mahachanok

Mahachanok is a hybrid mango variety developed in Thailand, combining characteristics of local and international cultivars. It is known for its slender, elongated shape and smooth, reddish-blushed yellow skin when ripe. The flesh is firm, fiberless, aromatic, and has a balanced taste—both sweet and slightly tangy. Mahachanok mangoes have excellent shelf life and transportability, making them a favorite for export markets and premium sales. (8).

2.1.1.2 Nam Dokmai Sithong

Nam Dokmai Sithong is one of Thailand’s most renowned premium mango varieties, prized for its elongated shape, golden-yellow skin, and rich fragrance. The flesh is soft, juicy, fiberless, and exceptionally sweet, with a delicate floral aroma. It is often enjoyed fresh or used in desserts like mango sticky rice. The variety matures earlier than many others and is highly valued for both domestic markets and export due to its consistent quality and elegant appearance (9).

2.1.1.3 R2E2

R2E2 is a large-sized mango variety originally developed in Australia but now widely grown in Thailand. It features a rounder, fuller shape compared to NDM and MHN, with bright, vibrant yellow-orange skin often overlaid with a red blush. The flesh is firm, slightly fibrous, mildly sweet, and less aromatic than Thai varieties, but it appeals to consumers who prefer a less intense flavor. Thanks to its size, attractive color, and good storage qualities, R2E2 is popular in both local and international markets (10).

2.1.2 Maturity

The maturity stage at which mangoes are harvested significantly affects their physicochemical properties, post-harvest handling, and overall quality. Maturity determination is critical for ensuring optimal flavor development, texture, and shelf life. The classification of mango maturity is based on physiological indicators such as peel color, firmness, total soluble solids (TSS), and starch degradation. Mangoes are typically categorized into three maturity stages:

2.1.2.1 Immaturity (M1)

Mangoes harvested at this stage are physiologically underdeveloped, with insufficient sugar accumulation and high acidity. They are primarily used for processing applications such as pickles, chutneys, and dried products rather than fresh consumption. Fruits harvested too early may fail to ripen properly and often exhibit inferior taste and texture.

2.1.2.2 Exporting Maturity (M2)

This stage represents mangoes that have reached physiological maturity but remain firm enough to endure transportation and storage. They typically possess an optimal balance between firmness and sugar content, allowing for controlled ripening during distribution. Harvesting mangoes at this stage ensures extended shelf life, making them suitable for export markets where delayed ripening is necessary to meet logistical and consumer demands.

2.1.2.3 Domestic Maturity (M3)

Fully ripened mangoes fall into this category, exhibiting peak sweetness, aroma, and soft texture. These mangoes are primarily sold in local markets for immediate consumption, as their advanced ripeness limits storage and transport duration. The sugar-to-acid ratio is at its highest, providing an ideal eating experience for consumers.

2.2 Related Work

Automated mango grading has advanced through machine learning and image processing, improving accuracy and efficiency. This study classifies NDM, MHN, and R2E2 mango varieties into three maturity stages (M1, M2, M3), leveraging these technologies. A web-based application is developed to support practical use, aiding decision-making in the mango supply chain.

Table 2.1 Qualitative comparison of related work

Paper	Dataset (No. of images)	Features	Classifier	Accuracy
(11)	Mangifera mango (900)	LAB, cross ratio, eccentricity, extent	Fuzzy systems	90
(12)	Harumanis mango (816)	RGB, contour detection	ANN	89
(13)	Alphonso mango (981)	LAB, region-based, contour-based	SVM	80
(14)	Alphonso mango (2400)	HSV, YUV, YCbCr	SVM, RF	97.6
(15)	Various mango (4010)	CNN feature extraction	CNN, SVM	83.16
(16)	Indian mango (1883)	MobileNet-v2, ShuffleNet	C-SVM	99.5

2.2.1 Mango Variety Detection and Classification

Advancements in mango type detection have also been significant. A study utilizing CNN-SVM models for mango type classification reported F1 scores between 83.16% and 85.99% (15). Additionally, in a more recent study, the classification of 15 Indian mango varieties was automated, achieving 99.5% accuracy through the combination of MobileNet-v2, ShuffleNet models, and a Cubic SVM classifier (16).

2.2.2 Mango Maturity Classification

Automated systems for grading mangoes have been extensively studied, with various machine learning and image processing techniques demonstrating promising results. In 2015, a study achieved 90% accuracy using fuzzy systems to classify mangoes based on shape, size, and maturity (11). A 2019 study employed image processing techniques for Harumanis mango maturity assessment, reaching 80% accuracy (12). In 2020, a hierarchical classification approach graded Alphonso mangoes into ripeness categories, achieving 88% accuracy using Support Vector Machines (SVM) (13). More recently, in 2024, research analyzing color features for ripeness classification achieved 97.6% accuracy using SVM and Random Forest classifiers (14). Additionally, a study in Bangladesh explored mango ripeness classification using various machine learning and deep learning models. Five classifiers—Gaussian Naive Bayes (GNB), SVM, Gradient Boosting (GB), Random Forest (RF), and K-Nearest Neighbors (KNN)—were tested alongside CNN and VGG16 for feature extraction. Results showed that CNN outperformed traditional methods, with Gradient Boosting achieving the highest accuracy of 96.28%, highlighting the effectiveness of deep learning in mango classification (17).

2.3 Theory of Computation

This section provides a fundamental exploration of the principles and capabilities of computational systems, which serve as a theoretical foundation for understanding and analyzing algorithms and computational models employed in this study.

2.3.1 VGG16

VGG16 is a deep convolutional neural network architecture known for its simplicity and effectiveness. The model consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. VGG16's design relies heavily on small (3x3) convolutional filters stacked on top of each other, which allows the model to learn complex features while keeping the architecture relatively simple. Despite its simplicity, VGG16 has been highly successful in many image classification tasks and has served as the foundation for many subsequent developments in deep learning. However, due to its large number of parameters, it can be computationally expensive to train, especially on large datasets (18).

2.3.2 Custom CNN or Lite VGG16

A Convolutional Neural Network (CNN) is a deep learning model specifically designed for processing image data. It mimics the structure of the human visual system by using convolutional layers to extract local features from images, followed by pooling layers that help reduce dimensionality (18). These features are then passed through fully connected layers to make predictions. CNNs are widely used in image classification, object detection, and semantic segmentation. The Custom CNN is a lightweight model for mango maturity classification, processing 128×128×3 RGB images with three convolutional layers (32, 64, 128 filters) and ReLU activation, followed by MaxPooling. After extracting features, a Flatten layer converts them into a 1D vector, and two dense layers (128 neurons, ReLU, softmax) provide the output. It uses the Adam optimizer, categorical cross-entropy loss, and is trained for 10 epochs with ImageDataGenerator for augmentation. This model, inspired by VGG16, is efficient and scalable, well-suited for automated mango maturity assessment, with potential improvements like dropout or batch normalization for enhanced performance.

2.3.3 ResNet50

ResNet50, a deep residual network, is designed to overcome the limitations of training very deep networks by using residual connections. These connections allow the network to skip certain layers, thereby mitigating the vanishing gradient problem and facilitating the training of networks with hundreds or even thousands of layers.

ResNet50 is known for its 50 layers and has been successful in achieving high performance on image classification tasks, particularly in challenges such as ImageNet. Its architecture enables the model to learn deep features without suffering from overfitting, making it highly effective for both small and large datasets (34).

2.3.4 InceptionV3

InceptionV3 is a deep convolutional neural network architecture that was developed as part of the GoogleNet project. One of its key innovations is the use of the "Inception module," which allows the model to learn multi-scale features by applying different types of convolution operations (e.g., 1x1, 3x3, 5x5) in parallel. This enables the network to capture both local and global features efficiently. InceptionV3 is known for its efficiency in terms of computational resources, achieving state-of-the-art results on large-scale image classification tasks while maintaining relatively low computational costs. It has become a go-to model for tasks that require high accuracy and fast inference times (20).

2.3.5 EfficientNet

EfficientNet is a family of deep neural networks that balance accuracy and efficiency by optimizing the depth, width, and resolution of the network through a compound scaling method. Unlike traditional CNN architectures, which increase depth or width separately, EfficientNet scales all dimensions together in a more balanced manner. As a result, it achieves higher accuracy with fewer parameters and lower computational costs. EfficientNet has become one of the most efficient models for image classification tasks, consistently outperforming other architectures in terms of both accuracy and efficiency on benchmark datasets (21).

2.3.6 Decision Tree

A Decision Tree (DT) is a machine learning algorithm that utilizes a tree structure to classify subjects based on an outcome. It does this by applying a splitting criterion, such as Gini impurity (Equation 1) or entropy (Equation 2), where ppp represents the probability of each class within a node. These criteria measure the impurity of a node. The algorithm recursively splits the data by evaluating all possible features and thresholds to identify the optimal split. This split is chosen to maximize information gain or minimize impurity, enabling the algorithm to effectively partition the data and make predictions based on learned patterns. Due to its intuitive nature, ease of understanding, high accuracy, and strong prediction capabilities, Decision Trees are widely used in prediction and classification tasks. In the context of mango variety classification, Decision Trees can be trained on various features extracted from mango images, such as shape, color, and texture, to distinguish between different mango varieties. The tree structure helps identify key characteristics that are most relevant for accurately classifying mango varieties, aiding in automated systems for sorting and quality control in agricultural practices (22).

$$Gini = 1 - \sum_{i=1}^{n} \left(P_i\right)^2 \tag{1}$$

$$Entropy = \sum_{i=1}^{n} - P_i log_2\left(P_i\right) \tag{2}$$

2.3.7 Random Forest

Random Forest (RF) is a powerful ensemble learning algorithm that enhances accuracy by combining multiple decision trees. It constructs individual trees using bootstrapped subsets of the training data and selects random feature subsets at each split. For classification tasks, RF determines the best split using either the Gini index (Equation 1) or entropy (Equation 2), while for regression, it minimizes variance (Equation 5). The final prediction is obtained through majority voting (Equation 3) for classification or averaging (Equation 4) for regression. Feature importance is assessed using Mean Decrease in Impurity (MDI), which quantifies the reduction in impurity across all splits where a feature appears. By leveraging multiple diverse decision trees, Random Forest effectively handles high-dimensional data, improves pattern recognition accuracy, and mitigates the risk of overfitting, structure shown in Figure 2.1.

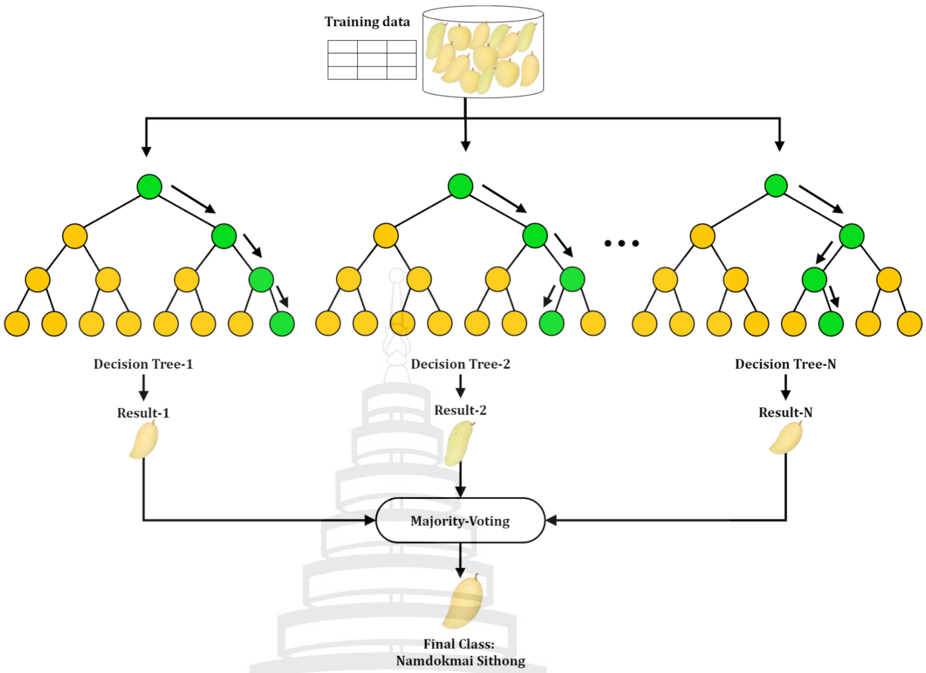


Figure 2.1 Illustration of random forest trees.

$$Majority\ Vote = \hat{y} = arg\ max_k \sum_{t=1}^{T} 1(y_t = k) \tag{3}$$

$$Averaging = \hat{y} = \frac{1}{T}\sum_{i=1}^{T} y_t \tag{4}$$

$$Variance = \frac{1}{n}\sum_{i=1}^{T} (yi - y)^2 \tag{5}$$

2.3.8 Gradient Boosting

Gradient Boosting (GB) is an ensemble learning technique that enhances prediction accuracy by combining multiple weak decision trees rather than relying on a single strong model. The algorithm iteratively minimizes errors by training decision trees to predict the residuals of the previous model, progressively refining predictions as the number of iterations increases and residuals decrease (25). It begins with an initial model and computes residuals by subtracting the actual values from the predicted ones (Equation 6). Weak learners are then trained on these residuals, and the model is updated by incorporating their weighted predictions (Equation 7), with the learning rate determining their contribution. The final prediction is obtained by aggregating the predictions of all weak learners, weighted by their respective learning rates (Equation 8). In the context of mango variety classification, Gradient Boosting can be used to improve the accuracy of predicting mango types based on features such as color, texture, and shape. By iteratively adjusting and combining predictions from multiple decision trees, the model can effectively differentiate between varieties like NDMi, MHN, and R2E2, making it a powerful tool for automated classification in agricultural applications.

$$Residual = Actual - Predicted \tag{6}$$

$$Model = Model + (Learning\ Rate * Weak\ Learner\ Prediction) \tag{7}$$

$$Prediction = \sum_{i=1}^{n} (Learning\ Rate * Weak\ Learner\ Predictions) \tag{8}$$

2.3.9 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a classification algorithm that assigns an instance based on the majority class of its k nearest neighbors in the feature space (26). It calculates distances using metrics like Euclidean (Equation 9) or Manhattan (Equation 10) and selects the k closest samples. Without explicitly defining class boundaries, KNN classifies mango varieties like NDM, MHN, and R2E2 based on shape, color, and texture, aiding automated sorting and quality control in agriculture.

$$Euclidean\ Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{9}$$

$$Manhattan\ Distance = |x_2 - x_1| + |y_2 - y_1| \tag{10}$$

2.3.10 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm primarily used for classification tasks. It aims to find the optimal hyperplane that separates data points of different classes with the maximum margin, meaning the greatest distance between the hyperplane and the nearest points from each class, known as support vectors. SVM focuses on maximizing this margin to improve the model’s generalization ability. When data is not linearly separable, SVM uses a technique called the kernel trick to project the data into a higher-dimensional space where a separating hyperplane can exist. By relying only on the support vectors, SVM creates a robust decision boundary that is less sensitive to outliers compared to other classifiers (27).

2.3.11 Color Value

Color feature extraction is a crucial aspect of image processing, particularly for tasks like image classification or object recognition. By analyzing the RGB, HSV, and LAB color spaces, we can extract statistical features such as mean (Equation 11), standard deviation (Equation 12), skewness (Equation 13), and kurtosis (Equation 14) to describe the color distribution in an image. In the RGB color space, the mean represents the average intensity of the red, green, and blue channels, while the standard deviation indicates color variation, and skewness and kurtosis reveal the symmetry and "tailedness" of pixel intensity distributions. In the HSV color space, the mean of the hue channel shows the dominant color, the saturation mean indicates color intensity, and the value mean reflects overall brightness, with skewness and kurtosis providing insights into color purity and brightness variations. The LAB color space, being perceptually uniform, uses the lightness channel to represent brightness and the A and B channels for color balance (green vs. red and blue vs. yellow), with statistical measures like mean, std, skewness, and kurtosis offering detailed insights into the image's color properties. These statistical features help to capture the diversity, asymmetry, and concentration of colors in an image, making them valuable for applications like image segmentation, classification, and recognition (28-30).

$$Mean = \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{11}$$

$$Standard\ Deviation = \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \tag{12}$$

$$Skewness = \gamma = \frac{1}{N\sigma^3}\sum_{i=1}^{N} (x_i - \mu)^3 \tag{13}$$

$$Kurtosis = k = \frac{1}{N\sigma^4}\sum_{i=1}^{N} (x_i - \mu)^4 \tag{14}$$

2.3.12 Texture Value

Haralick features are extracted from the gray-level co-occurrence matrix (GLCM), which represents how frequently pairs of pixel with specific values occur in an image. Haralick features capture texture patterns such as contrast (Equation 15), correlation, energy, and homogeneity (31). These features are essential for analyzing the surface patterns of objects in images, like the skin of mangoes, to differentiate between varieties. Local Binary Pattern (LBP) (Equation 16) is a texture descriptor used to characterize the local structure of an image. It works by comparing the intensity of each pixel to its neighboring pixels and assigning binary values based on this comparison (32). These binary patterns are then encoded into a histogram to represent the texture. LBP is effective in capturing fine-grained textures and has been widely used in facial recognition and object classification tasks.

$$Contrast = \sum_{i=0}^{P-1}\sum_{j=0}^{P-1}(i - j)^2 P(i, j) \quad (15)$$

$$LBP(x_c) = \sum_{p=0}^{P-1} s(g_p - g_c)^{2^p} \quad (16)$$

2.3.13 Shape Value

Contour-based shape feature extraction is a key technique in computer vision that characterizes objects by analyzing their contours and outlines, providing essential geometric and morphological properties. This research categorizes shape feature extraction into three main types: Geometric Features, Morphological Features, and Circularity and Moments. Geometric features describe the object's physical dimensions, including area, perimeter, and aspect ratio, where area represents the number of pixels inside the contour, perimeter measures boundary length, and aspect ratio determines elongation. Morphological features evaluate shape structure, compactness, and convexity, including compactness, extent, solidity, and eccentricity, which describe how closely an object approximates a circle and how concave it is. Circularity and moments help identify roundness and shape invariance, where circularity quantifies how close an object is to a perfect circle, and Hu moments provide shape descriptors invariant to scaling, rotation, and translation, making them essential for pattern recognition. These extracted features play a vital role in applications such as fruit classification, object detection, and quality assessment in computer vision (33).

$$Area(A) = \sum_{i=4}^{N} 1 \tag{17}$$

$$Perimeter(P) = \sum_{i=4}^{N} d(i) \tag{18}$$

$$Aspect\ Ratio(AR) = \frac{W}{H} \tag{19}$$

$$Compactness(C) = \frac{P^2}{A} \tag{20}$$

$$Extent(E) = \frac{A}{Abbox} \tag{21}$$

$$Solidity(S) = \frac{A}{Ahull} \tag{22}$$

$$Eccentricity(EC) = \sqrt{1 - \frac{b^2}{a^2}} \tag{23}$$

$$Circularity(CIR) = \frac{4\pi A}{P^2} \tag{24}$$

$$Hu\ Moments(H) = function\ of\ \mu ij \tag{25}$$

N (Total Pixel Count) – Represents the total number of pixels enclosed within the contour, corresponding to the area occupied by the object in the image.

M (Number of Contour Points) – Denotes the total number of discrete points that define the boundary of the contour.

$d(i)$ (Euclidean Distance Between Consecutive Points) – Represents the Euclidean distance between successive contour points, which characterizes the smoothness and complexity of the boundary.

W (Bounding Rectangle Width) – Defines the horizontal extent of the smallest axis-aligned rectangle that completely encloses the contour..

H (Bounding Rectangle Height) – Defines the vertical extent of the smallest axis-aligned rectangle that fully contains the contour.

$Ahull$ (Convex Hull Area) – Represents the area of the convex hull, which is the smallest convex shape that can entirely enclose the contour.

$Abbox$ (Bounding Box Area) – Denotes the area of the minimal upright bounding rectangle, computed as $Abbox = W \times H$.

2.3.14 Evaluation Metrics

Evaluation metrics are quantitative measures used to assess a machine learning model's performance and predictive capabilities. In this study, five metrics were used to evaluate and compare model performance.

2.3.14.1 Accuracy

Accuracy is a ratio of correctly predictive results of observations to total observations, as shown in Equation (26).

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{26}$$

2.3.14.2 Precision

Precision is the proportion of correctly predicted positive observations to all predicted positive observations, as shown in Equation (27).

$$Precision = \frac{TP}{TP+FP} \tag{27}$$

2.3.14.3 Recall

Recall or Sensitivity is the proportion of correctly predicted positive observations to all observations in the positive of the actual class as in Equation (28).

$$Recall = \frac{TP}{TP+FN} \tag{28}$$

2.3.14.4 F1 Score

F1 Score is the weighted average calculation of the precision and recall value as in Equation (29), to provides a balanced measure of the model’s performance.

$$F1\,score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{29}$$

CHAPTER 3

RESEARCH METHODOLOGY

The research methodology of the study is illustrated, covering data collection, data preprocessing, feature engineering, and model construction.

3.1 Research Overview

This research presents a structured methodology for mango variety classification and maturity assessment using machine learning and deep learning techniques, consisting of five main stages: data collection, data preparation, variety classification, maturity classification, and web-based implementation. Data was collected in collaboration with local farmers and the School of Agro-Industry, resulting in two dataset versions—Version 1 (1,544 images) for variety classification and Version 2 (1,432 images) for both variety and maturity classification. The preprocessing pipeline involved background removal, image resizing, and data augmentation to enhance model performance. For variety classification, both dataset versions were used with different training-validation-test splits and feature extraction techniques, applying machine learning models such as RF, SVM, and KNN to classify mangoes into three varieties: NDM, MHN, and R2E2. For maturity classification, only Dataset Version 2 was used, categorizing mangoes into ripening stages (M1, M2, M3 for NDM & MHN; M1, M2 for R2E2), with both traditional machine learning models and deep learning (CNN-based) approaches applied for evaluation. To enable real-time classification, a web application was developed, integrating a Next.js frontend with a Flask API backend, allowing users to upload mango images and receive predictions for both variety and maturity stage. This methodology offers an efficient, scalable, and non-destructive solution for agricultural applications, contributing to advancements in precision farming and supply chain optimization, as illustrated in Figure 3.1.

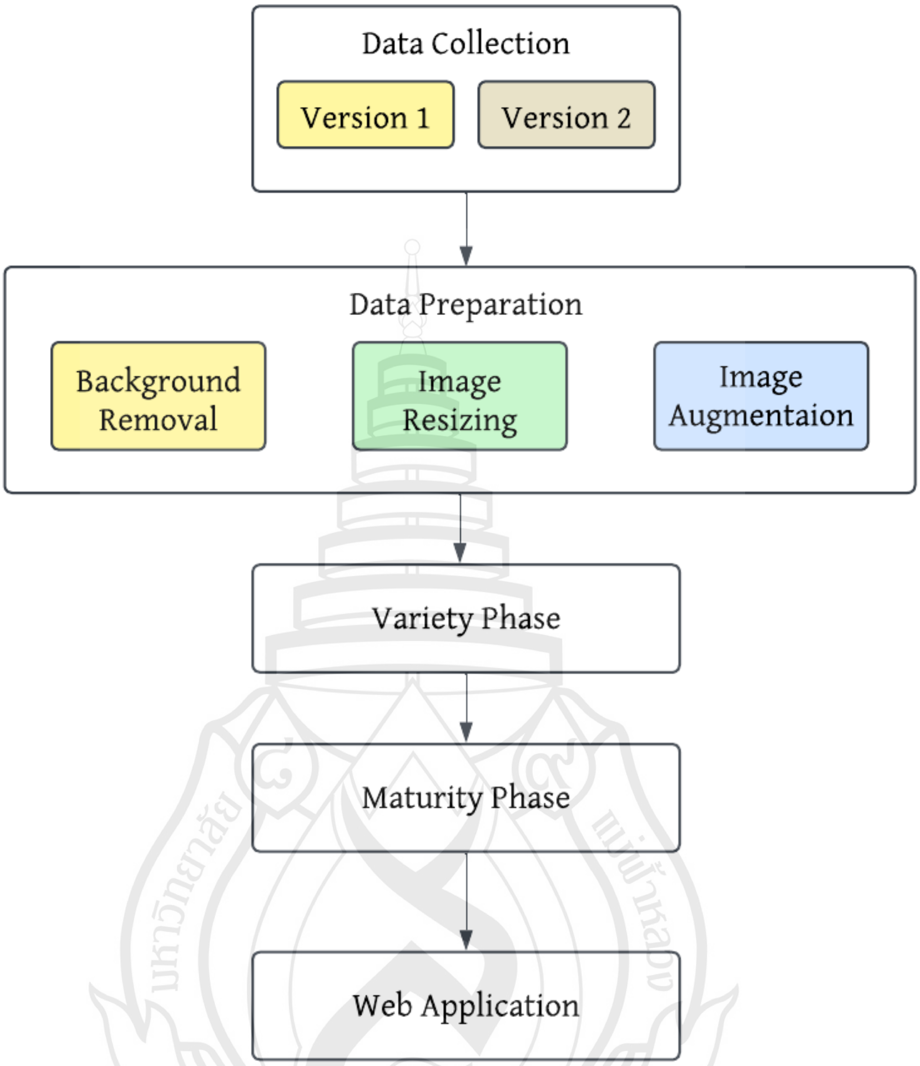


Figure 3.1 Overview of the Methodology Process

3.2 Data Collection and Preparation

The data collection phase was meticulously designed, resulting in two curated versions of the dataset for mango variety and maturity classification. Version 1 comprises 1,544 mango samples (Table 3.1) dedicated to variety classification, consisting of 600 MHN, (38.9%), 500 NDM, (32.4%), and 444 R2E2 (28.7%), split into 80% for training, 10% for testing, and 10% for validation to ensure robust model development. Version 2 includes 1,432 mango samples (Table 3.2) designed for both variety and maturity classification, with data partitioned into 60% training, 20% testing, and 20% validation for variety classification, and an 80% training and 20% testing split for maturity classification per variety example on Table 3.3. To maintain high dataset quality, a collaborative process was established between local farmers and the School of Agro-Industry at Mae Fah Luang University, where farmers initially classified the mangoes by variety and maturity stage before forwarding them to the university for standardized high-resolution imaging under controlled conditions. For maturity classification, farmers employed the water and salt flotation method, a non-destructive technique based on fruit buoyancy in salt solutions (5%–15%), where immature mangoes sink in 5%, mature green sink in 10% but float in lower concentrations, breaking stage float in 10% but sink in 15%, and ripe mangoes float in 15%; this method, leveraging density variations due to changes in water content during ripening, was further validated using supplementary indicators such as peel color, firmness, Brix level, acidity, and aroma to ensure classification accuracy. As shown in Table 3.4, The dataset preparation phase further enhanced data quality through an image processing workflow consisting of background removal to isolate the mango fruit from irrelevant visual noise, image resizing to standardize input dimensions, and data augmentation to increase dataset diversity and improve model robustness, ensuring that the images used for feature extraction and model training were of consistently high quality. By integrating a systematic data collection strategy, validated classification methods, and a structured dataset preparation pipeline, this study provides a high-quality dataset that enhances machine learning model performance.

Table 3.1 Dataset Specifications for Three Commercial Mango Varieties (Version 1)

Aspect	Description
Specific subject area	Mango image classification using machine learning and deep learning
Mango varieties	NDM, MHN, R2E2
Image data type	Digital images (JPG), RGB color space, max width 640px
Data acquisition	This research examines three mango varieties that play a significant role in Thailand's mango market. Data was gathered by individually photographing each mango against a white background in the Agro-Industry School's laboratory.
Number of images	A total of 1,544 images were collected including 600 of NDM, 500 of MHN, and 444 of R2E2.
Dataset accessibility	https://github.com/xzodus000/mango-3-class-max-w-500.git

Table 3.2 Dataset Specifications for Three Commercial Mango Varieties (Version 2)

Aspect	Description
Specific subject area	Mango image classification using machine learning and deep learning
Mango varieties	NDM, MHN, R2E2
Mango maturity	M1, M2, M3
Image data type	Digital images (JPG), RGB color space, max width 640px
Data acquisition	This research examines three mango varieties that play a significant role in Thailand's mango market. Data was gathered by individually photographing each mango against a white background in the Agro-Industry School's laboratory.
Number of images	A total of 1,432 images were collected including

	640 of NDM, 372 of MHN, and 420 of R2E2.
Dataset accessibility	https://github.com/xzodus000/dataset-mfu-maturity-mango.git

Table 3.3 Dataset Example

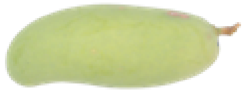
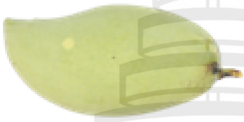
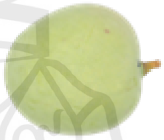
	M1	M2	M3
MHN			
NDM			
R2E2			

Table 3.4 Image Data Preparation Results

Original	Background Removal & Resizing	Image Augmentation
		 0° Rotation  90° Rotation

3.3 Variety Phase

In the mango variety classification phase, two dataset versions were utilized, each employing distinct data splitting strategies and feature extraction techniques to optimize classification performance. Dataset Version 1 was divided into 80% training, 10% validation, and 10% testing and incorporated color features from RGB, HSV, and LAB color spaces, texture features extracted using Local Binary Pattern (LBP) and Haralick descriptors, and shape-based features, including contour analysis and eccentricity. Classification models such as Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were implemented, with performance evaluated using K-fold cross-validation and confusion matrices as shown in Figure 3.2. In contrast, Dataset Version 2 employed a 60% training, 20% validation, and 20% testing split and included an expanded set of shape-related descriptors, such as morphological features, geometric features, circularity, and moments, while retaining the same color and texture features as in Dataset Version 1. The same classification models were applied, and model robustness was assessed using K-fold cross-validation and confusion matrix analysis. The comparative evaluation of these datasets provided insights into the influence of feature extraction techniques and data distribution on classification accuracy, emphasizing the strengths and limitations of each approach in distinguishing mango varieties, as summarized in Figure 3.3.

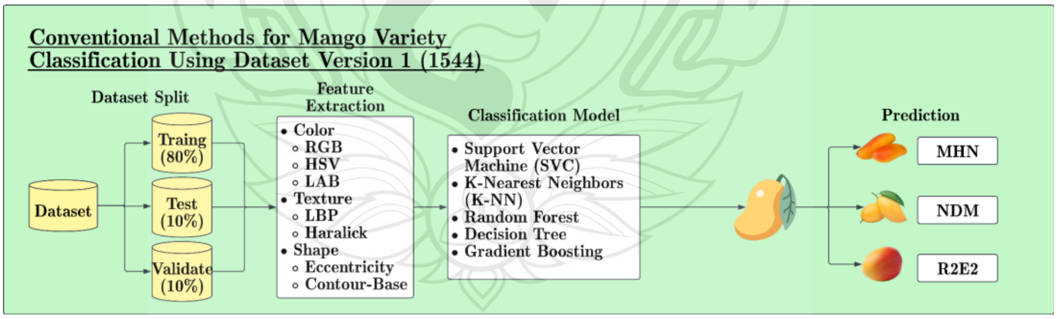


Figure 3.2 Conventional Methods for Mango Variety Classification Using Dataset Version 1

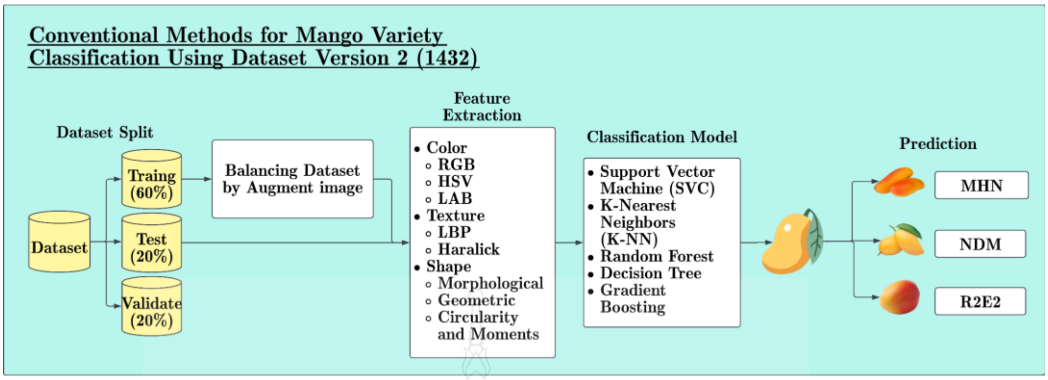


Figure 3.3 Conventional Methods for Mango Variety Classification Using Dataset Version 2

3.4 Maturity Phase

In the mango maturity classification phase, both machine learning and deep learning approaches were applied using Dataset Version 2 to assess the maturity stages of mangoes. In the machine learning-based classification, the dataset was split into 80% training and 20% validation, with feature extraction focusing on color features from RGB, HSV, and LAB color spaces, as well as texture features derived from Local Binary Pattern (LBP) and Haralick descriptors. Classification models, including Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), were implemented, with performance evaluated using K-fold cross-validation and confusion matrices. Meanwhile, in the deep learning-based classification, the dataset was similarly divided into 80% training and 20% validation, and pre-trained convolutional neural networks (CNNs), including VGG16, Lite VGG16, ResNet50, InceptionV3, and EfficientNet, were employed to extract high-level features for classification. Model performance was assessed using K-fold cross-validation and confusion matrix analysis. This dual approach allowed for a comparative analysis of traditional feature-based machine learning models versus deep learning architectures in determining mango maturity stages, as summarized in Figure 3.4.

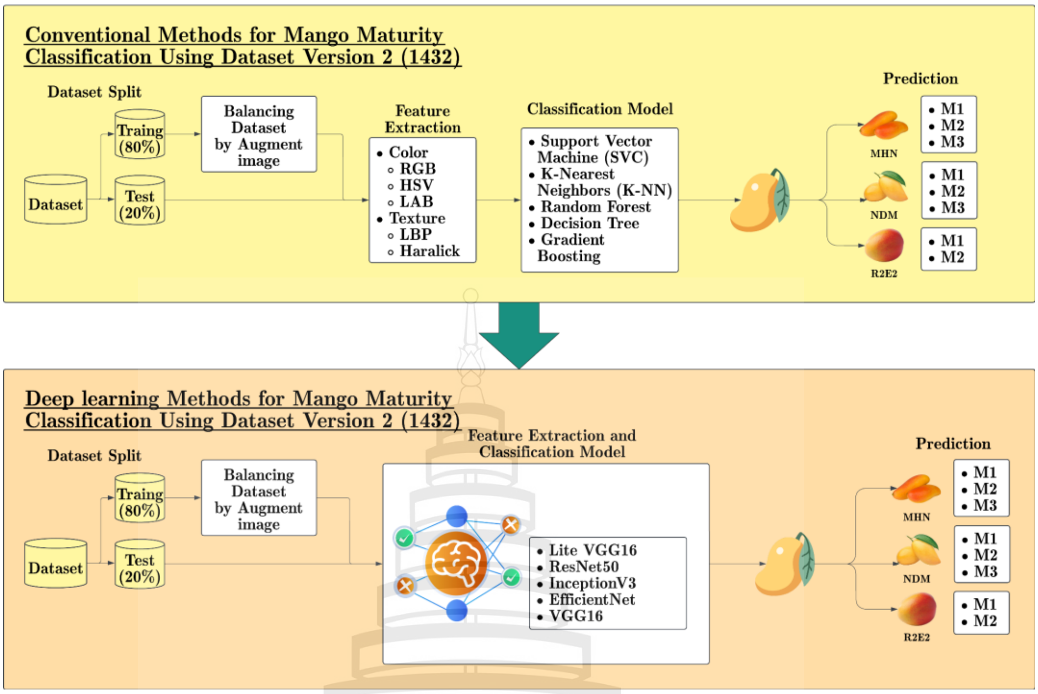


Figure 3.4 Methodology Process for Mango Maturity Phase Classification

3.5 Developing a Web Application

A web-based application was developed for real-time mango variety and maturity classification, with a frontend built using Next.js for an interactive and responsive user interface. The backend, implemented with a Flask API, handles image processing and model inference. Upon receiving an image, the system performs feature extraction to analyze the mango's characteristics, such as color, shape, and texture, and then classifies the mango variety (e.g., MHN, NDM, R2E2). If only the variety classification is required, the backend returns the predicted mango variety. However, if the predicted variety is MHN, the system uses the MHN mango maturity model to classify the maturity stage of the MHNvariety. Similarly, for NDM, the system applies the NDM mango maturity model for maturity classification, and for R2E2, the R2E2 mango maturity model is used. This automated, non-destructive solution supports practical applications in agriculture and the supply chain, providing both mango variety classification and maturity assessment as shown in Figure 3.5.

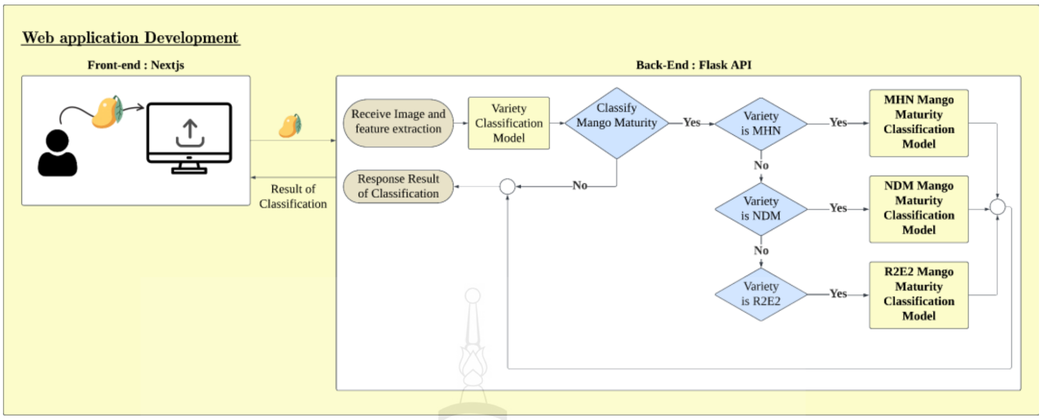


Figure 3.5 Web Application Architecture

CHAPTER 4

EXPERIMENTAL RESULTS

This section presents the experimental outcomes for classifying three Thai mango varieties (NDM, MHN, and R2E2) and their maturity stages (M1, M2, M3). The data collection phase was meticulously designed, resulting in two dataset versions to support both variety and maturity classification. Version 1 consists of 1,544 mango samples, used for variety classification. Version 2, comprising 1,432 mango samples, was used for both variety and maturity classification. The dataset was analyzed using shape, color, and texture features. Various classifiers, including Decision Trees, Random Forest, Gradient Boosting, Support Vector Machines, and K-Nearest Neighbors, were optimized for performance. A 5-fold cross-validation strategy was employed to enhance robustness. Through a structured data collection and validation process, combined with systematic dataset design, this study provides a reliable foundation for high-performance machine learning models supporting non-destructive postharvest assessment and agricultural automation.

4.1 Comparison for mango varieties classification with dataset version 1

This experiment presents the classification results of three mango varieties — NDM, MHN, and R2E2 — using Dataset Version 1. Feature extraction was performed based on color spaces, shape and texture characteristics to support classification tasks. Five machine learning models, including Decision Tree, RF, GB, KNN, and SVM, were evaluated with an 80-10-10 split for training, testing, and validation.

4.1.1 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification

The experimental results for mango variety classification using dataset version 1 based on color feature extraction (RGB, HSV, and LAB) with various classifiers are presented in Table 4.1. Among all classifiers, the Random Forest (RF) classifier demonstrated the highest accuracy of 92.90% when using HSV color features, followed closely by 90.32% accuracy in both LAB and RGB color spaces. Gradient Boosting (GB) also showed strong and consistent performance, achieving 90.32% accuracy with LAB features and slightly lower results in RGB (89.67%) and HSV (89.03%). Decision Trees (DT) performed notably well, with the highest accuracy of 90.96% achieved when utilizing LAB color features. The K-Nearest Neighbors (KNN) classifier produced moderate results, with HSV features giving the best accuracy at 85.80%. In contrast, the Support Vector Machine (SVM) classifier demonstrated relatively poor performance across all color spaces, with its best result being 53.40% using RGB features and lower accuracies in HSV (49.84%) and LAB (60.84%). Overall, using dataset version 1, the findings indicate that ensemble classifiers, particularly Random Forest and Gradient Boosting, combined with color feature extraction in HSV and LAB color spaces, provide the most effective and reliable performance for mango variety classification.

Table 4.1 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification (Dataset Version 1)

Classifier	**Accuracy**		
	RGB	**HSV**	**LAB**
DT	83.87%	83.87%	90.96%
RF	89.67%	**92.90%**	90.32%
GB	89.67%	89.03%	90.32%
KNN	83.87%	85.80%	82.58%
SVM	53.40%	49.84%	6084%

4.1.2 Comparison of Classifier Performance Based on Shape Feature Extraction for Mango Variety Classification

In Table 4.2, The performance of various machine learning classifiers was assessed for mango variety classification using shape feature extraction methods from Dataset Version 1, including Contour-Based, Eccentricity, and a combination of both features (Contour-Based & Eccentricity). Random Forest (RF) demonstrated the highest classification accuracy, achieving 98.00% with the combined feature set, highlighting its robustness in handling multiple features. Decision Tree (DT) and

Gradient Boosting (GB) followed closely, both reaching an accuracy of 97.06% with the combined feature set, indicating their strong performance. K-Nearest Neighbors (KNN) showed a lower accuracy of 92.90% with the combined feature set, while still performing well, but less effectively than the aforementioned classifiers. Support Vector Machine (SVM) underperformed, achieving a maximum accuracy of 85.16% with the combined feature set, suggesting it was less suited for this task. In conclusion, Random Forest was the most effective model for mango variety classification based on shape features in Dataset Version 1, while SVM showed the least effectiveness compared to the other classifiers.

Table 4.2 Comparison of Classifier Performance Based on Shape Feature Extraction for Mango Variety Classification (Dataset Version 1)

Classifier	**Accuracy**		
	Contour-Base	**Eccentricity**	**Contour-Base & Eccentricity**
DT	86.45%	95.48%	97.06%
RF	87.09%	96.12%	**98.00%**
GB	86.45%	95.48%	97.06%
KNN	86.45%	96.08%	92.90%
SVM	47.57%	79.27%	85.16%

4.1.3 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification

This research conducted a comparative analysis of two texture feature extraction techniques, Local Binary Patterns (LBP) and Haralick texture features, to assess their effectiveness in classifying mango species based on skin texture analysis. The objective was to determine which feature extraction strategy provided better performance and discrimination for mango species classification. The results, presented in Table 4.3, showed that LBP features generally outperformed Haralick features across various machine learning classifiers. The Random Forest classifier, in particular, demonstrated exceptional performance, achieving the highest accuracy of 97.37% when using LBP features. Additionally, it performed well when both texture feature sets were applied, showcasing its ability to handle diverse texture variations. These findings suggest that LBP features, which focus on local textural patterns, are well-suited for identifying mango species. The study highlights the importance of selecting appropriate feature extraction methods and classifiers to optimize classification accuracy in mango species identification, emphasizing the potential of LBP features and machine learning classifiers in achieving high accuracy for automated mango sorting and classification systems.

Table 4.3 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification (Dataset Version 1)

Classifier	**Accuracy**	
	Haralick	**LBP**
DT	92.75%	92.68%
RF	95.79%	**97.37%**
GB	88.73%	**97.37%**
KNN	66.65%	95.93%
SVM	41.19%	87.80%

4.1.4 Comparison of Classifier Performance Based on Combine Feature Extraction for Mango Variety Classification

In Table 4.4, The experimental results for mango variety classification using combined feature extraction techniques — including HSV color features, contour-based and eccentricity shape features, and LBP texture descriptors — demonstrated that ensemble classifiers outperformed other models. Among the five classifiers evaluated on Dataset Version 1, the RF classifier achieved the highest performance with an accuracy of 98.06%, precision of 98.09%, recall of 98.06%, and an F1-score of 98.06%. GB followed closely with consistent scores of 97.56% across all metrics. The DT classifier showed moderate performance, with accuracy, precision, recall, and F1-score all at 76.77%. On the other hand, the KNN classifier delivered lower results, with accuracy at 60.70%, precision at 61.40%, recall at 60.70%, and an F1-score of 60.82%. The Support Vector Machine (SVM) recorded the weakest performance, with an accuracy of 53.33%, precision of 37.30%, recall of 53.33%, and an F1-score of 43.15%. These results indicate that the integration of color, shape, and texture features is highly effective for mango variety classification, particularly when using ensemble learning models such as Random Forest and Gradient Boosting.

Table 4.4 Comparison of Classifier Performance Based on Combined Features Extraction for Mango Variety Classification (Dataset Version 1)

Classifier	Accuracy			
	Accuracy	**Precision**	**Recall**	**F1-score**
DT	76.77%	76.77%	76.77%	76.77%
RF	**98.06%**	**98.09%**	**98.06%**	**98.06%**
GB	97.56%	97.56%	97.56%	97.56%
KNN	60.70%	61.40%	60.70%	60.82%
SVM	53.33%	37.30%	53.33%	43.15%

4.1.5 Hyperparameter Tuning of Random Forest Classifier for Mango Variety Classification (Dataset Version 1)

In Table 4.4, the RF model achieves an impressive accuracy of 98.06% in Figures 4.1, underscoring the importance of hyperparameter tuning to enhance performance. The hyperparameter optimization of the RF classifier, which incorporated a combination of HSV color features, contour-based shape descriptors, eccentricity, and LBP texture features, significantly improved its ability to distinguish between mango varieties. The tuning process utilized an advanced grid search to fine-tune key parameters, including maximum depth, maximum number of leaf nodes, and the number of estimators. The parameter values explored were: maximum depth set to None, 3, 10, and 20; a maximum of 9 leaf nodes; and the number of estimators set at 150 and 200. The optimal configuration, which included a maximum depth of 3, 9 leaf nodes, and 200 estimators, resulted in a remarkable accuracy of 98.71% in Figures 4.2. This optimization achieved the highest classification accuracy for mango varieties, leveraging the combined features of HSV color, contour-based shape, eccentricity, and LBP texture. The confusion matrix presented in Figure 4.2 further illustrates the RF model's performance after hyperparameter tuning, highlighting its outstanding accuracy of 98.71%. The model correctly classified 53 images of NDM, 43 images of MHN, and 57 images of R2E2, while also identifying some similarities between NDM and MHN, as well as between R2E2 and NDM. Although a few misclassifications were observed, the model demonstrated excellent precision and recall, with minimal false positives and negatives, as depicted in Figures 4.3.

Prediction Class	Actual Class: MHN	NDM	R2E2	Precision
MHN	53	1	1	96.36%
NDM	0	43	0	100%
R2E2	0	0	57	100%
Recall	100%	97.73%	98.28%	Accuracy: 98.71%

Figure 4.1 Confusion Matrix RF Mango Variety Before Hyperparameter Tuning (Dataset Version 1)

Prediction Class	Actual Class: MHN	NDM	R2E2	Precision
MHN	53	1	1	96.36%
NDM	0	43	0	100%
R2E2	0	0	57	100%
Recall	100%	97.73%	98.28%	Accuracy: 98.06%

Figure 4.2 Confusion Matrix RF Mango Variety After Hyperparameter Tuning (Dataset Version 1)

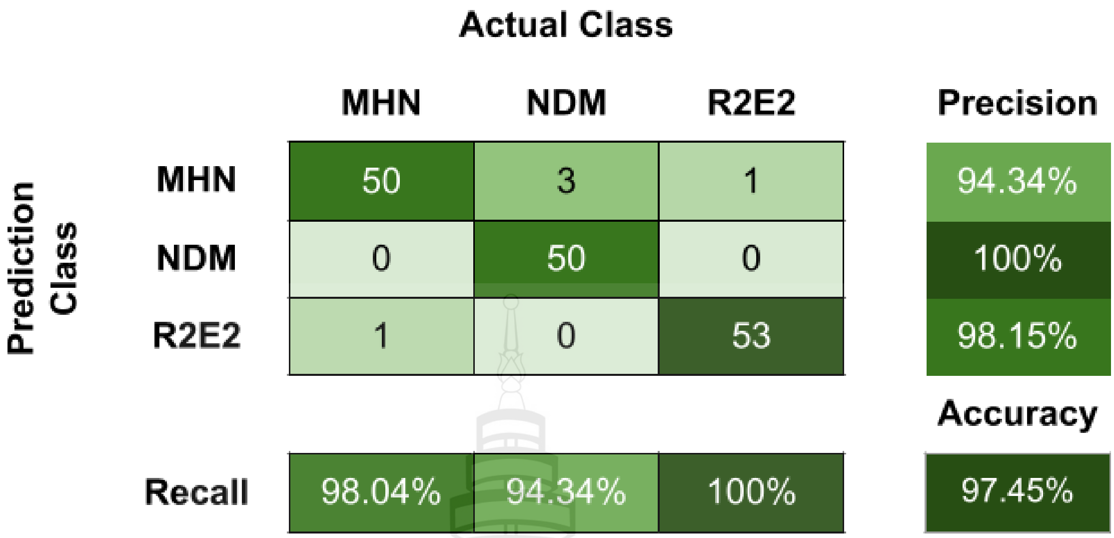


Figure 4.3 Confusion Matrix of Random Forest Model for Mango Variety Classification After Hyperparameter Tuning with Unseen Data (Dataset Version 1)

4.2 Comparison for Mango Varieties Classification with Dataset Version 2

This experiment investigates the classification of three mango varieties — NDM, MHN, and R2E2 — utilizing Dataset Version 2. Feature extraction was conducted using color information from the RGB, HSV, and LAB color spaces, in addition to shape descriptors based on contour analysis, and texture features derived from Haralick and LBP methods. Five machine learning models — Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) — were assessed using a 60-20-20 data split for training, testing, and validation.

4.2.1 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification

Table 4.5 presents the comparative performance of various classifiers based on color feature extraction methods (RGB, HSV, and LAB) for mango variety classification using Dataset Version 2. The results indicate that all classifiers performed exceptionally well, with accuracies exceeding 96% across all color spaces. The Decision Tree (DT) classifier achieved the highest accuracy of 97.75% using HSV features, followed closely by its performance with RGB and LAB features at 97.37%

and 97.74%, respectively. The K-Nearest Neighbors (KNN) classifier also showed strong results, attaining 97.74% accuracy with HSV and 97.37% with both RGB and LAB. Both Random Forest (RF) and Gradient Boosting (GB) classifiers demonstrated consistent and robust performance with identical accuracies of 97.00% across all three color spaces. Meanwhile, the Support Vector Machine (SVM) classifier achieved 97.00% accuracy with RGB and HSV and slightly lower accuracy of 96.62% with LAB. Overall, these findings highlight the effectiveness and reliability of color-based feature extraction methods for mango variety classification, with HSV features providing a slight advantage for certain classifiers.

Table 4.5 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification (Dataset Version 2)

Classifier	**Accuracy**		
	RGB	**HSV**	**LAB**
DT	97.37%	97.75%	97.74%
RF	97.00%	97.00%	97.00%
GB	97.00%	97.00%	97.00%
KNN	97.37%	97.74%	97.37%
SVM	97.00%	97.00%	96.62%

4.2.2 Comparison of Classifier Performance Based on Shape Feature Extraction for Mango Variety Classification

This experiment investigates the extraction of various shape features from mango images to enhance the classification of different mango varieties. Circularity/Moments, Morphological, and Geometric features were utilized to provide valuable insights into the morphological variations of mangoes. The results, as presented in Table 4.6, demonstrate that Morphological features achieved the highest classification accuracy, with models consistently exhibiting strong performance in terms of precision, recall, and overall classification. Specifically, Morphological features, such as compactness and solidity, achieved an accuracy of 98.13% when the Random Forest Classifier was employed, emphasizing their effectiveness in distinguishing between mango varieties. The integration of these shape features significantly enhances classification accuracy, offering comprehensive insights into mango morphology, which is crucial for advancements in both agricultural practices and computer vision applications. In the context of mango maturity classification, shape features have proven particularly effective, delivering unmatched accuracy and detailed information. As such, the scope of this research does not include experiments focused on the use of shape features for mango maturity classification.

Table 4.6 Comparison of Classifier Performance Based on Shape Feature Extraction for Mango Variety Classification (Dataset Version 2)

Classifier	**Accuracy**		
	Circularity Moments	**Morphological**	**Geometric**
DT	95.88%	97.38%	95.51%
RF	97.38%	**98.13%**	96.25%
GB	95.88%	98.13%	95.51%
KNN	96.63%	97.75%	82.39%
SVM	91.76%	97.75%	79.40%

4.2.3 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification

The results of the experiment, as shown in Table 4.7, demonstrate the comparative performance of Haralick and Local Binary Pattern (LBP) texture features for classifying mango varieties using various machine learning classifiers. LBP features consistently outperformed Haralick features across all classifiers. Notably, the Gradient Boosting (GB) classifier achieved the highest accuracy of 97.75% with LBP features, followed by Random Forest (RF) at 96.28%, and Decision Trees (DT) at 93.63%. K-Nearest Neighbors (KNN) also showed improved accuracy with LBP (92.13%) compared to Haralick (66.65%), while Support Vector Machine (SVM) demonstrated the lowest accuracy, with 73.03% for LBP and 41.19% for Haralick. These results highlight the superior ability of LBP features to capture fine-grained textural patterns, making it a more effective choice for mango variety classification.

Table 4.7 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification (Dataset Version 2)

Classifier	**Accuracy**	
	Haralick	**LBP**
DT	92.75%	93.63%
RF	95.79%	96.28%
GB	88.73%	**97.75%**
KNN	66.65%	92.13%
SVM	41.19%	73.03%

4.2.4 Comparison of Classifier Performance Based on Combine Feature Extraction for Mango Variety Classification

In Table 4.8, The experimental results of combined feature extraction for mango variety classification, using features from HSV color space, morphological, and Local Binary Pattern (LBP) texture, were evaluated using Dataset Version 2. The results reveal that morphological features alone outperformed all classifiers with an outstanding accuracy of 98.13%. RF followed closely with an accuracy of 97.82%, indicating its strong performance, while GB achieved a 97.65% accuracy, showcasing its reliability. KNN performed well with a 96.16% accuracy, and DT also yielded a respectable 95.28% accuracy. On the other hand, SVM showed significantly lower performance with an accuracy of 76.09%. These results suggest that the morphological feature set is highly effective for mango variety classification, and ensemble methods like RF and GB are strong contenders, although the morphological features alone provide superior performance in this task.

Table 4.8 Comparison of Classifier Performance Based on Combined Features Extraction for Mango Variety Classification (Dataset Version 2)

Classifier	Accuracy			
	Accuracy	**Precision**	**Recall**	**F1-score**
DT	95.28%	95.30%	95.28%	95.26%
RF	**97.82%**	**97.82%**	**97.82%**	**97.82%**
GB	97.65%	97.65%	97.65%	97.65%
KNN	96.16%	96.19%	96.16%	96.16%
SVM	76.09%	75.25%	76.09%	75.29%

4.2.5 Hyperparameter Tuning of Random Forest Classifier for Mango Variety Classification (Dataset Version 2)

In Table 4.6, the RF model achieves an impressive accuracy of 98.06%, The hyperparameter tuning of the Random Forest classifier using morphological features for mango variety classification demonstrated substantial improvements in model performance. Key parameters, including max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators, were optimized through a grid search approach. The optimal configuration was identified as follows: max_depth set to None, max_features set to the square root of the total number of features, min_samples_leaf set to 1, min_samples_split set to 2, and n_estimators set to 200. With these optimized parameters, the classifier achieved a cross-validated accuracy of 99.25% and an overall test accuracy of 99.25% as In Figure 4.5. Furthermore, when evaluated on unseen validation data, the classifier exhibited even stronger performance, reaching an accuracy of 99.63%. However, it was observed that incorporating additional color features for classifying NDM and R2E2 varieties did not result in further improvements, as the model’s performance remained consistent without these features. as depicted in Figures 4.4 to 4.6.

Prediction Class \ Actual Class	MHN	NDM	R2E2	Precision
MHN	67	4	0	94.37%
NDM	1	107	0	99.07%
R2E2	0	0	88	100%
				Accuracy
Recall	98.53%	96.40%	100%	98.12%

Figure 4.4 Confusion Matrix RF Mango Variety Before Hyperparameter Tuning (Dataset Version 2)

Prediction Class \ Actual Class	MHN	NDM	R2E2	Precision
MHN	67	1	0	98.53%
NDM	1	110	0	99.10%
R2E2	0	0	88	100%
				Accuracy
Recall	98.53%	99.10%	100%	99.25%

Figure 4.5 Confusion Matrix RF Mango Variety After Hyperparameter Tuning (Dataset Version 2)

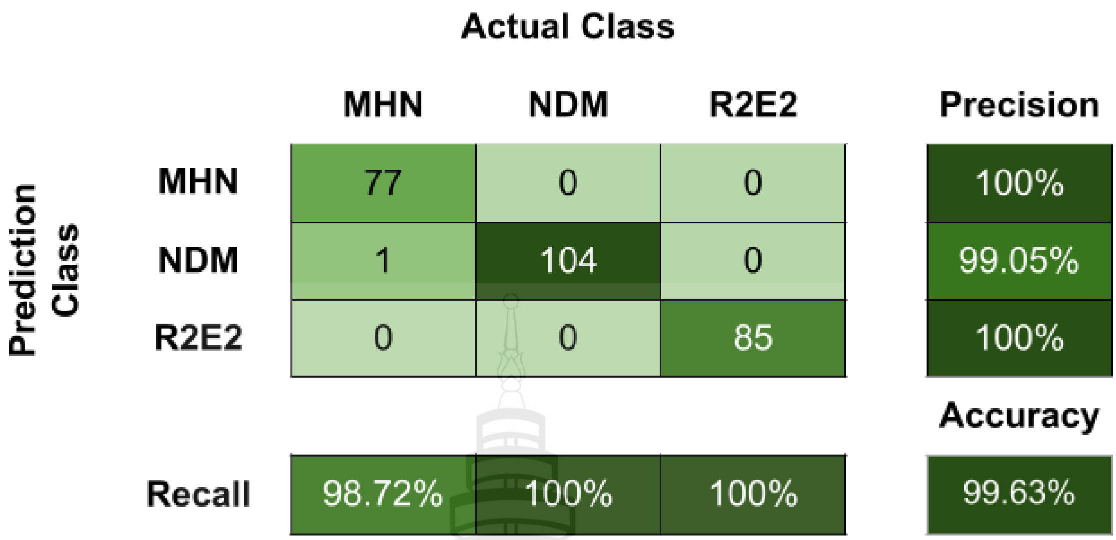


Figure 4.6 Confusion Matrix RF Mango Variety After Hyperparameter Tuning with Unseen Data (Dataset Version 2)

4.3 Comparison of Mango Maturity Classification Using Dataset Version 2 for MHN Mango

This experiment investigates the assessment of maturity levels in MHN mangoes using Dataset Version 2. Feature extraction was conducted using color information from the RGB, HSV, and LAB color spaces, in addition to shape descriptors based on contour analysis, and texture features derived from Haralick and LBP methods. Five machine learning models — DT, RF, GB, KNN, and SVM — were assessed using a 80, 20 data split for training, testing.

4.3.1 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification

In Table 4.9, The experimental results for mango maturity classification (MHN variety) based on color feature extraction revealed varying classifier performances across different color spaces. The Decision Tree (DT) classifier achieved the highest accuracy with RGB features at 56.00%, followed by HSV at 58.67%, and LAB at 54.67%. The Random Forest (RF) classifier performed similarly, with an accuracy of 58.67% for RGB, but dropped to 53.33% for both HSV and LAB. Gradient Boosting (GB) showed the best performance with RGB features at 60.00%, but its accuracy declined to 45.33% with HSV and 50.67% with LAB. K-Nearest Neighbors (KNN) achieved the lowest accuracy for RGB features at 48.00%, while its performance was slightly better with LAB (54.67%) and HSV (49.33%). Support Vector Machine (SVM) consistently underperformed across all color spaces, with a maximum accuracy of 30.67%. These results demonstrate the influence of color space choice on classifier performance for mango maturity classification, with RGB generally yielding better outcomes.

Table 4.9 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Maturity MHN Mango Classification

Classifier	Accuracy		
	RGB	**HSV**	**LAB**
DT	56.00%	58.67%	54.67%
RF	58.67%	53.33%	53.33%
GB	**60.00%**	45.33%	50.67%
KNN	48.00%	49.33%	54.67%
SVM	30.67%	30.67%	30.67%

4.3.2 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification

The results of the experiment, as shown in Table 4.10, demonstrate the comparative performance of Haralick and Local Binary Pattern (LBP) texture features for classifying mango varieties using various machine learning classifiers. LBP features consistently outperformed Haralick features across all classifiers. Notably, the Gradient Boosting (GB) classifier achieved the highest accuracy of 97.75% with LBP features, followed by Random Forest (RF) at 96.28%, and Decision Trees (DT) at 93.63%. K-Nearest Neighbors (KNN) also showed improved accuracy with LBP (92.13%) compared to Haralick (66.65%), while Support Vector Machine (SVM) demonstrated the lowest accuracy, with 73.03% for LBP and 41.19% for Haralick. These results highlight the superior ability of LBP features to capture fine-grained textural patterns, making it a more effective choice for mango variety classification.

Table 4.10 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Maturity MHN Mango Classification

Classifier	Accuracy	
	Haralick	LBP
DT	56.00%	52.00%
RF	54.67%	57.33%
GB	54.67%	**60.00%**
KNN	25.33%	52.00%
SVM	30.67%	30.67%

4.3.3 Comparison of Deep Learning Models for Mango Maturity Classification

The performance evaluation of deep learning models for MHN mango maturity classification, as presented in Table 4.11, highlights varying levels of effectiveness across different architectures and training epochs. Among the tested models, InceptionV3 achieved the highest accuracy of 80.00% over 9 epochs, demonstrating strong feature extraction capabilities, while Lite VGG16 followed with 77.26% accuracy at 10 epochs, indicating its efficiency despite being a lightweight variant. VGG16 and ResNet50 recorded moderate accuracies of 63.21% and 54.85%, respectively, with ResNet50's lower performance possibly attributed to its deeper architecture requiring more training iterations. In contrast, EfficientNet exhibited the lowest accuracy at 38.13% over 9 epochs, suggesting its feature extraction approach might not be well-suited for this classification task. The results indicate that InceptionV3 and Lite VGG16 are the most effective models for MHN mango maturity classification, emphasizing the importance of choosing an appropriate architecture and training duration to optimize performance.

Table 4.11 Performance of Deep Learning in MHN Mango Maturity Classification

Classifier	Train Accuracy	Accuracy	Loss	Epochs
Lite VGG16	79.60%	77.26%	22.74%	10
ResNet50	60.54%	54.85%	45.15%	8
InceptionV3	**83.95%**	**80.00%**	**20.00%**	**9**
EfficientNet	46.15%	38.13%	61.87%	9
VGG16	69.57%	63.21%	36.79%	10

4.4 Comparison of Mango Maturity Classification Using Dataset Version 2 for NDM Mango

This experiment investigates the assessment of maturity levels in NDM mangoes using Dataset Version 2. Feature extraction was conducted using color information from the RGB, HSV, and LAB color spaces, in addition to shape descriptors based on contour analysis, and texture features derived from Haralick and LBP methods. Five machine learning models — DT, RF, GB, KNN, and SVM — were assessed using a 80, 20 data split for training, testing.

4.4.1 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification

The experimental results for mango maturity classification (NDM variety) based on color feature extraction demonstrated varying classifier performances across different color spaces, as shown in Table 4.12. The Decision Tree (DT) classifier achieved the highest accuracy with LAB features at 84.40%, followed closely by RGB at 83.49%, while HSV performed lower at 71.56%. Random Forest (RF) performed well with RGB (81.65%) and LAB (83.49%) features, while its accuracy decreased with HSV at 79.82%. Gradient Boosting (GB) showed the best performance with RGB at 84.40%, but its accuracy dropped with HSV (72.48%) and LAB (81.65%). K-Nearest Neighbors (KNN) achieved the lowest accuracy with HSV at 55.96%, while its performance was better with RGB (70.64%) and LAB (66.97%). Support Vector Machine (SVM) consistently underperformed across all color spaces, with accuracy values ranging from 55.05% to 55.96%. These results indicate that LAB and RGB color features are more effective for classifying mango maturity in the NDM variety, with DT and GB classifiers yielding the highest accuracies.

Table 4.12 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Maturity NDM Mango Classification

Classifier	**Accuracy**		
	RGB	**HSV**	**LAB**
DT	83.49%	71.56%	**84.40%**
RF	81.65%	79.82%	83.49%
GB	**84.40%**	72.48%	81.65%
KNN	70.64%	55.96%	66.97%
SVM	55.05%	55.05%	55.96%

4.4.2 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification

The Haralick and Local Binary Pattern (LBP) texture features were evaluated for classifying mango maturity in the MHN variety, with the results presented in Table 4.13. The findings reveal that LBP features generally outperformed Haralick features across various machine learning classifiers. Notably, the Random Forest classifier achieved the highest accuracy of 57.33% using LBP features, effectively capturing local texture patterns. This highlights the potential of LBP features in improving mango maturity classification accuracy. The results suggest that LBP features are particularly effective at capturing fine-grained textural details, leading to better classification performance. Accuracy results for classifiers, such as Decision Trees (56.00% with LBP), Random Forest (54.67% with LBP), Gradient Boosting (54.67% with LBP), K-Nearest Neighbors (25.33% with LBP), and Support Vector Machine (30.67% with LBP), emphasize the importance of selecting the appropriate feature extraction techniques and classifiers for improved classification accuracy. These findings indicate that optimizing both feature extraction methods and classifier choices can lead to significant advancements in mango maturity classification.

Table 4.13 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Maturity NDM Mango Classification

Classifier	**Accuracy**	
	Haralick	**LBP**
DT	66.05%	52.00%
RF	**71.56%**	57.33%
GB	62.38%	60.00%
KNN	58.72%	52.00%
SVM	55.96%	30.67%

4.4.3 Comparison of Deep Learning Models for Mango Maturity Classification

The performance evaluation of deep learning models for NDM mango maturity classification, as shown in Table 4.14, demonstrates varying levels of accuracy across different architectures and training epochs. InceptionV3 achieved the highest accuracy of 82.46% over 10 epochs, indicating its strong feature extraction capabilities for this classification task. VGG16 and Lite VGG16 followed with accuracies of 65.89% and 60.62%, respectively, both trained for 10 epochs, suggesting that these architectures effectively capture relevant features with sufficient training. ResNet50 recorded a lower accuracy of 55.17% despite also being trained for 10 epochs, which may indicate challenges in optimizing its deep architecture for this dataset. EfficientNet showed the lowest accuracy at 38.79%, completing training in just 2 epochs, implying that insufficient training may have hindered its performance. Overall, the results highlight InceptionV3 as the most effective model for NDM mango maturity classification, with VGG16 and Lite VGG16 also demonstrating moderate effectiveness, emphasizing the importance of architecture selection and training duration in optimizing classification performance.

Table 4.14 Performance of Deep Learning in NDM Mango Maturity Classification

Classifier	Train Accuracy	Accuracy	Loss	Epochs
Lite VGG16	64.91%	60.62%	39.38%	10
ResNet50	53.22%	55.17%	44.83%	10
InceptionV3	**86.94%**	**82.46%**	**17.54%**	**10**
EfficientNet	47.95%	38.79%	61.21%	2
VGG16	67.45%	65.89%	34.11%	10

4.5 Comparison of Mango Maturity Classification Using Dataset Version 2 for R2E2 Mango

This experiment investigates the assessment of maturity levels in R2E2 mangoes using Dataset Version 2. Feature extraction was conducted using color information from the RGB, HSV, and LAB color spaces, in addition to shape descriptors based on contour analysis, and texture features derived from Haralick and LBP methods. Five machine learning models — DT, RF, GB, KNN, and SVM — were assessed using a 80, 20 data split for training, testing.

4.5.1 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Variety Classification

The experimental results for mango maturity classification (R2E2 variety) based on color feature extraction, as shown in Table 4.15, revealed varying performances across different classifiers and color spaces. The Decision Tree (DT) classifier achieved the highest accuracy with RGB and LAB features, both at 79.76%, while its performance dropped to 75.00% with HSV. Random Forest (RF) performed well with RGB at 83.33%, followed closely by LAB at 82.14%, and 75.00% with HSV. Gradient Boosting (GB) showed similar results, with an accuracy of 79.76% for RGB, 75.00% for HSV, and 78.57% for LAB. K-Nearest Neighbors (KNN) achieved the lowest accuracy with RGB at 65.48%, but performed better with HSV (73.81%) and LAB (67.86%). Support Vector Machine (SVM) consistently underperformed across all color spaces, with the highest accuracy of 67.86% for RGB and the lowest for LAB at 57.14%. These results indicate that RGB and LAB features provide better classification performance for the R2E2 variety, with Random Forest and Decision Tree classifiers yielding the highest accuracies.

Table 4.15 Comparison of Classifier Performance Based on Color Feature Extraction for Mango Maturity R2E2 Mango Classification

Classifier	**Accuracy**		
	RGB	**HSV**	**LAB**
DT	79.76%	75.00%	79.76%
RF	**83.33%**	75.00%	82.14%
GB	79.76%	75.00%	78.57%
KNN	65.48%	73.81%	67.86%
SVM	67.86%	64.29%	57.14%

4.5.2 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Variety Classification

The results from Table 4.16 show the comparison of Haralick and LBP texture features for classifying mango maturity in the R2E2 variety. LBP features consistently outperformed Haralick features across all classifiers. The RF classifier achieved the highest accuracy of 80.95% with LBP, followed by GB at 78.57%. DT had an accuracy of 77.38% with LBP, while KNN and SVM showed lower performance, especially with Haralick features. These findings suggest that LBP features offer a significant advantage in classifying mango maturity for the R2E2 variety, enhancing classification accuracy.

Table 4.16 Comparison of Classifier Performance Based on Texture Feature Extraction for Mango Maturity R2E2 Mango Classification

Classifier	**Accuracy**	
	Haralick	**LBP**
DT	61.90%	77.38%
RF	72.62%	**80.95%**
GB	70.24%	78.57%
KNN	64.29%	76.19%
SVM	64.29%	57.14%

4.5.3 Comparison of Deep Learning Models for Mango Maturity Classification

The performance evaluation of deep learning models for R2E2 mango maturity classification, as presented in Table 4.17, demonstrates varying classification accuracies across different architectures and training epochs. InceptionV3 achieved the highest accuracy of 82.49% over 10 epochs, confirming its strong feature extraction capabilities for this classification task. Lite VGG16 followed with 78.34% accuracy in 9 epochs, while VGG16 and ResNet50 attained 74.78% and 70.62% accuracy, respectively, over 8 and 10 epochs, indicating their effectiveness in capturing relevant maturity features. EfficientNet recorded the lowest accuracy at 56.68% despite being trained for 10 epochs, suggesting its feature extraction approach may not be as well-suited for R2E2 mango classification. Overall, the results highlight InceptionV3 as the most effective model, with Lite VGG16, VGG16, and ResNet50 also demonstrating strong performance, emphasizing the impact of architecture choice and training duration on classification accuracy.

Table 4.17 Performance of Deep Learning in R2E2 Mango Maturity Classification

Classifier	Train Accuracy	Accuracy	Loss	Epochs
Lite VGG16	64.09%	60.62%	39.38%	10
ResNet50	61.13%	55.17%	44.83%	10
InceptionV3	**84.87%**	**82.46%**	**17.54%**	**10**
EfficientNet	46.29%	38.79%	61.21%	2
VGG16	70.03%	65.89%	34.11%	10

4.6 Model Evaluation Results

This section examines the process of fine-tuning hyperparameters in the Random Forest model employed for the classification of Thai mangoes. A 5-fold cross-validation was used to evaluate model performance in preparation for further performance enhancement in real-world scenarios.

4.6.1 Comparison with Validation Results

Table 4.18 presents the results of the 5-fold cross-validation performed on Dataset Version 1 using the Random Forest classifier with combined shape, color, and texture features. The model demonstrated consistently high accuracy across all folds, with accuracies ranging from 98.25% to 99.56%. The highest accuracy of 99.56% was achieved in both the fourth and fifth folds, while the third fold recorded the lowest accuracy at 98.25%. Furthermore, the mean squared error (MSE) remained constant at 0.44 across all folds, indicating stable and reliable model performance. These results highlight the robustness and effectiveness of the Random Forest model in classifying mango varieties when utilizing integrated features

Table 4.18 5-Fold Cross-Validation Results for Mango Variety Classification Using Random Forest (Dataset Version 1)

Fold	Accuracy	Loss (MSE)
1	99.12	0.44
2	98.68	0.44
3	98.25	0.44
4	99.56	0.44
5	99.56	0.44

Table 4.19 displays the results of the 5-fold cross-validation using Dataset Version 2 with the Random Forest classifier based on combined features. The model continued to perform with high consistency and accuracy, ranging from 97.38% to 99.62%. The highest accuracy, 99.62%, was achieved in the fourth fold, while the third fold showed the lowest accuracy at 97.38%. Similar to Dataset Version 1, the mean squared error (MSE) remained constant at 0.44 across all folds, reflecting the model's stable prediction capability. These results further confirm the reliability and robustness of the Random Forest classifier in accurately identifying mango varieties when applied to different dataset versions.

Table 4.19 5-Fold Cross-Validation Results for Mango Variety Classification Using Random Forest (Dataset Version 2)

Fold	Accuracy	Loss (MSE)
1	98.13	0.44
2	99.25	0.44
3	97.38	0.44
4	99.62	0.44
5	99.25	0.44

Table 4.20 presents the results of 5-fold cross-validation for MHN mango maturity classification using the InceptionV3 deep learning model on Dataset Version 2. The model achieved accuracies ranging from 77.00% to 80.00%, with the highest accuracy recorded in folds 2 and 5. The mean squared error (MSE) values ranged from 0.60 to 0.66, reflecting the model's consistent but moderate performance. Compared to machine learning approaches, the InceptionV3 model demonstrated reliable predictive capability, though there remains room for improvement in achieving higher accuracy for maturity classification tasks.

Table 4.20 5-Fold Cross-Validation Results for MHN Mango Maturity Classification Using InceptionV3 (Dataset Version 2)

Fold	Accuracy	Loss (MSE)
1	78.00	0.65
2	80.00	0.60
3	77.00	0.66
4	79.00	0.62
5	80.00	0.60

Table 4.21 presents the results of 5-fold cross-validation for NDM mango maturity classification using the GB classifier on Dataset Version 2. The model achieved accuracies ranging from 71.30% in fold 3 to a maximum of 84.40% in fold 1. The corresponding loss (MSE) values ranged between 0.16 and 0.29, showing moderate variability across folds. These results indicate that the Gradient Boosting model performed well in classifying NDM mango maturity levels, though with slightly fluctuating performance between folds.

Table 4.21 5-Fold Cross-Validation Results for NDM Mango Maturity Classification Using Gradient Boosting (Dataset Version 2)

Fold	Accuracy	Loss (MSE)
1	84.40	0.16
2	76.15	0.24
3	71.30	0.29
4	76.85	0.23
5	81.48	0.18

Table 4.22 presents the 5-fold cross-validation results for R2E2 mango maturity classification using the Random Forest classifier on Dataset Version 2. The model demonstrated good overall performance, with accuracies ranging from 72.62% in fold 5 to a peak of 82.14% in fold 1. The mean squared error (MSE) varied between 0.18 and 0.27 across the folds. These results suggest that while the model performed reliably, there was slightly lower consistency in later folds, indicating potential data variability or complexity in R2E2 maturity classification.

Table 4.22 5-Fold Cross-Validation Results for R2E2 Mango Maturity Classification Using Random Forest (Dataset Version 2)

Fold	Accuracy	Loss (MSE)
1	82.14	0.18
2	78.57	0.21
3	80.95	0.19
4	78.57	0.21
5	72.62	0.27

Next, the Random Forest model was evaluated using a separate 20% of the dataset, which was reserved for testing. The model achieved an impressive overall accuracy of 99.63% in classifying mango variety, effectively distinguishing between MHN, NDM, and R2E2 varieties, as shown in Figure 4.6. The model correctly classified 104 instances of NDM, 77 instances of MHN, and 85 instances of R2E2, with only a few misclassifications.

Precision and recall metrics further validated the model's performance, with precision rates of 99.05% for NDM, and 100% for both MHN and R2E2, indicating minimal false positives. Recall rates of 100%, 98.72%, and 100% for NDM, MHN, and R2E2, respectively, demonstrated the model’s ability to capture nearly all instances for each class, with minimal false negatives. These results highlight the robustness of the Random Forest model and its potential for real-world applications, such as mango farming and marketing, where accurate classification is crucial for ensuring product quality and market competitiveness.

Prediction Class \ Actual Class	MHN	NDM	R2E2	Precision
MHN	8	12	0	40.00%
NDM	2	38	0	95.00%
R2E2	3	1	14	93.33%
				Accuracy
Recall	80.00%	74.51%	100%	80.00%

Figure 4.7 Confusion Matrix for InceptionV3 Model in Mango Maturity Classification (MHN)

Prediction Class \ Actual Class	MHN	NDM	R2E2	Precision
MHN	54	2	0	96.43%
NDM	6	19	3	67.86%
R2E2	0	6	19	76.00%
				Accuracy
Recall	90.00%	70.37%	86.36%	84.40%

Figure 4.8 Confusion Matrix for GB Model in Mango Maturity Classification (NDM)

Prediction Class \ Actual Class	MHN	NDM	Precision
MHN	54	2	96.43%
NDM	6	19	67.86%
Recall	90.00%	70.37%	Accuracy: 83.33%

Figure 4.9 Confusion Matrix for RF Model in Mango Maturity Classification (R2E2)

Table 4.23 Summary of System Accuracy

Variety	Variety Accuracy	Maturity Accuracy
MHN	99.63%	80.52%
NDM	99.63%	85.58%
R2E2	99.63%	84.71%

The mango classification system showcases exceptional performance, making it a reliable tool for both identifying mango varieties and assessing their maturity levels. The system achieves a remarkable 99.63% accuracy in classifying mango varieties, with this high accuracy consistently observed across the three key mango types: Mahachanok (MHN), Nam Dokmai Sithong (NDM), and R2E2. This impressive result underscores the system's ability to reliably differentiate between mango varieties, a critical task in the agricultural industry where accurate classification plays a vital role in sorting, packaging, and exporting mangoes. Beyond variety classification, the system also excels in maturity assessment, with accuracies of 80.52% for MHN, 85.58% for NDM, and 84.71% for R2E2, reflecting its robustness in recognizing ripeness stages. Although maturity classification is inherently more complex due to the gradual and subtle changes in fruit characteristics as it ripens, the system's performance remains strong, demonstrating its capability to handle diverse classification challenges. The system’s success in both variety and maturity classification proves its practical value in agricultural applications, particularly for tasks like determining optimal harvest times, reducing postharvest losses, and ensuring the quality of mangoes throughout the supply chain. Its high accuracy levels are a testament to the integration of advanced machine learning and computer vision techniques, making it a game-changing tool for the mango industry, particularly in regions with large-scale mango production and export needs.

4.7 Web Application

Figure 4.10 Web user scenario for mango variety classification

The user journey on the Mango Variety Classification website starts when the user lands on the homepage, where they are welcomed with an overview of the tool’s purpose and a “Start Inspection” button. Upon clicking the button to begin, the user uploads a clear image of a mango, choosing from supported formats like JPG or PNG. this is link of web mango classification http://iate-mango-classification.com The website processes the uploaded image by extracting key features. Once the image has been processed, the system classifies the mango variety (MHN, NDM, or R2E2) and displays the result along with additional relevant details.

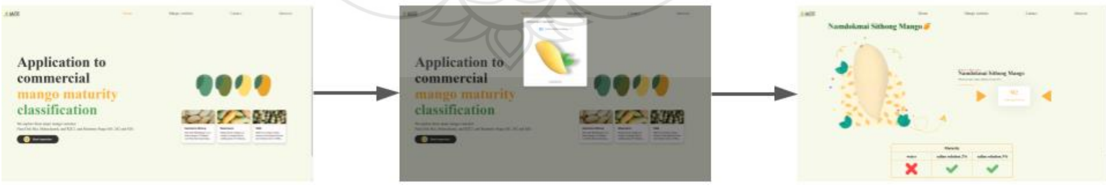


Figure 4.11 Web user scenario for mango maturity classification

The user journey on the Mango Maturity Classification website begins when the user arrives on the homepage, greeted with an introduction to the tool's purpose and a "Start Inspection" button. After clicking the button to begin, the user uploads a clear image of a mango in an accepted format, such as JPG or PNG. Before classifying the mango's maturity, the system first identifies the mango variety. Based on the variety prediction, the corresponding maturity classification model is selected. The website then processes the image by extracting relevant features related to the mango's maturity. Once the processing is complete, the system classifies the mango's maturity level (M1, M2, M3) and displays the result, along with additional details such as suggested harvest timing or ripening information.

The table 4.24 shows the runtime performance of the web-based mango classification system for both variety and maturity detection. Variety classification is the fastest, taking just 2.11 seconds to process. For maturity detection, the runtimes vary by mango type: MHN maturity detection takes 6.38 seconds, NDM is quicker at 4.15 seconds, and R2E2 maturity detection completes in 6.22 seconds. These results indicate an overall efficient system, with slight variations in processing time depending on the specific maturity classification task.

Table 4.24 Performance Timing (s) for Mango Classification by Variety and Maturity Stages

Category	**Runtime (seconds)**
Variety	2.11s
MHN Maturity	6.38s
NDM Maturity	4.15s
R2E2 Maturity	6.22s

4.8 Discussion

The MHN maturity classification model encounters challenges in accurately distinguishing between different maturity stages, largely due to limitations in the dataset, which may not fully capture variations across diverse environments and growth stages. Additionally, using separate models for mango variety and maturity increases computational costs and reduces efficiency. The use of two versions of the dataset — one for variety classification and another for maturity classification — was initially intended to optimize performance for each specific task, as variety classification relies more on physical attributes like shape and texture, while maturity classification focuses on color changes and surface details. However, this approach introduces complexity in data management and model training, requiring the maintenance and processing of separate datasets, which increases workload and raises the risk of inconsistencies. To address these challenges, future work will aim to develop an integrated model capable of classifying both mango variety and maturity stages simultaneously, thereby improving efficiency and accuracy. Expanding the dataset with a broader range of samples and employing advanced machine learning techniques, such as transfer learning and multi-task learning, will further enhance the model’s performance and reliability.

CHAPTER 5

CONCLUSIONS

This research successfully developed and evaluated machine learning models for mango variety and maturity classification. After hyperparameter tuning, Dataset Version 2 proved to be the most effective for mango variety classification, with the Random Forest classifier achieving a cross-validated accuracy of 99.25% and an impressive 99.63% on unseen validation data, outperforming Dataset Version 1, which achieved 98.71% with combined HSV color, shape, and LBP texture features. The higher accuracy of Dataset Version 2, despite using only morphological features, suggests that this feature set is more optimal for variety classification. The highest maturity classification accuracies were 80.00% for Mahachanok using InceptionV3, 84.40% for Namdokmai Sithong using Gradient Boosting, and 83.33% for R2E2 using Random Forest. respectively, highlighting the potential for accurately predicting mango maturity stages (M1, M2, M3). The integration of optimized feature extraction methods and ensemble learning models has proven highly effective for both variety and maturity classification. The deployment of these models in a user-friendly web application showcases the practical potential of this research, providing a tool for real-time postharvest quality control and agricultural automation, ultimately benefiting mango producers and exporters.decision-making, postharvest quality control, and agricultural automation for mango producers and exporters.

REFERENCES

1 Laguerre O, Duret S, Chaomuang N, Denis A, Derens-Bertheau E, Mawilai P, et al. Influence of long-distance air transport conditions on horticultural product quality: Case study of fresh mango shipment from Thailand to France. Int J Refrig. 2023;152:16–25. doi:10.1016/j.ijrefrig.2023.04.023

2 FAOStat, FAO, Others. Food and agriculture data. Crop Statistics [Internet]. 2020 [cited 2024 May 19]. Available from: http://www.fao.org/faostat

3 Tharanathan RN, Yashoda HM, Prabha TN. Mango (Mangifera indica L.), "The King of Fruits"—An Overview. Food Rev Int. 2006;22(2):95–123. doi:10.1080/87559120600574493

4 Food and Agriculture Organization of the United Nations. Statistics | FAO [Internet]. [cited 2024 May 19]. Available from: https://www.fao.org/statistics/en

5 Statista. Thailand: monthly fresh and frozen mango export volume 2024 [Internet]. [cited 2024 May 19]. Available from: https://www.statista.com/statistics/1319653/thailand-monthly-fresh-and-frozen-mango-export-volume/

6 Department of Agricultural Extension. Agricultural Production Information Service System [Internet]. [cited 2024 May 19]. Available from: https://production.doae.go.th/service/data-state-product/index

7 Kitinoja L, Kader A. Measuring postharvest losses in fruits and vegetables in developing countries. 2015. doi:10.13140/RG.2.1.3921.6402

8 Vayo Group. Mahachanok Mango or the Royal Favorite (Mangifera indica L.) [Internet]. 2024 [cited 2024 May 19]. Available from: https://farm.vayo.co.th/blog/mangifera-indica-l-mahachanok-mango/

9 Vayo Group. Nam Dokmai Sithong Mango (Mangifera indica L.) [Internet]. 2024 [cited 2024 May 19]. Available from: https://farm.vayo.co.th/blog/mangifera-indica-l-mahachanok-mango/

10 Vayo Group. R2E2 Mango (Mangifera indica) [Internet]. 2024 Jan 22 [cited 2024 May 19]. Available from: https://farm.vayo.co.th/blog/r2e2-mangifera-indica/

11 Sudheer KP, Indira V. Post Harvest Technology of Horticultural Crops [Internet]. [cited 2024 May 19]. Available from: https://books.google.co.th/books?id=Ws_SQUB5tj0C

12 Naik S, Patel B, Pandey R. Shape, size and maturity features extraction with fuzzy classifier for non-destructive mango (Mangifera indica L., cv. Kesar) grading. 2015 IEEE TIAR. 2015:1–7. doi:10.1109/TIAR.2015.7358522

13 Ahmad KAB, Othman M, Syed Abdullah SL, Rasidah Ali N, Muhamat Dawam SR. Mango Shape Maturity Classification Using Image Processing. 2019 4th ICCRAIE. 2019:1–5. doi:10.1109/ICRAIE47735.2019.9037776

14 Raghavendra A, Guru DS, Rao MK, Sumithra R. Hierarchical approach for ripeness grading of mangoes. Artif Intell Agric. 2020;4:243–52. doi:10.1016/j.aiia.2020.10.003

15 Prabhu A, Rani NS, Nandi CS. Towards importance of comprehensive color features analysis using iterative golden ratio proportions for Alphonso mango ripening stage classification by adapting to natural progressive ripening method. J Food Compos Anal. 2024;126:105873. doi:10.1016/j.jfca.2023.105873

16 Yashu, Kukreja V, Madan K. Automated Mango Type Classification Using CNN-SVM Ensemble: A Deep Learning Approach for Quality Control and Market Analysis in the Mango Industry. 2023 ICCCIS. 2023:261–5. doi:10.1109/ICCCIS60361.2023.10425736

17 Ratha AK, Barpanda NK, Sethy PK, Behera SK. Automated classification of Indian mango varieties using machine learning and MobileNet-v2 deep features. Trait Signal. 2024;41(2):669–78. doi:10.18280/ts.410210

18 Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. 2017 ICET. 2017:1–6. doi:10.1109/icengtechnol.2017.8308186

20 Cao J, Yan M, Jia Y, Tian X, Zhang Z. Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals. EURASIP J Adv Signal Process. 2021;2021(1). doi:10.1186/s13634-021-00740-8

21 Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019. doi:10.48550/ARXIV.1905.11946

22 Huynh-Cam TT, Chen LS, Le H. Using Decision Trees and Random Forest Algorithms to Predict and Determine Factors Contributing to First-Year University Students' Learning Performance. Algorithms. 2021;14(11):318. doi:10.3390/a14110318

23 Guo X, Hao P. Using a Random Forest Model to Predict the Location of Potential Damage on Asphalt Pavement. Appl Sci. 2021;11(21):10396. doi:10.3390/app112110396

24 Purwanto AD, Wikantika K, Deliar A, Darmawan S. Decision Tree and Random Forest Classification Algorithms for Mangrove Forest Mapping in Sembilang National Park, Indonesia. Remote Sens. 2022;15(1):16. doi:10.3390/rs15010016

25 Wang D, Hwang J, Lee J, Kim M, Lee I. Temperature-Based State-of-Charge Estimation Using Neural Networks, Gradient Boosting Machine and a Jetson Nano Device for Batteries. Energies. 2023;16(6):2639. doi:10.3390/en16062639

26 Gao X, Li G. A KNN Model Based on Manhattan Distance to Identify the SNARE Proteins. IEEE Access. 2020;8:112922–31. doi:10.1109/access.2020.3003086

27 Pisner DA, Schnyer DM. Support vector machine. In: Machine Learning. 1st ed. Cambridge: Academic Press; 2020. p. 101–21. doi:10.1016/b978-0-12-815739-8.00006-7

28 Chen W, Fubin W, Xianzhong C, Qingwen H, Zhikun C, Meng Z, et al. RGB color decomposition and image feature extraction of flame image in rear of sintering machine. 2017 CCC. 2017:5460–3. doi:10.23919/ChiCC.2017.8028222

29 Ma JQ. Content-Based Image Retrieval with HSV Color Space and Texture Features. 2009 WISM. 2009:61–3. doi:10.1109/WISM.2009.20

30 Barbu T, Ciobanu A, Costin M. Unsupervised Color-based Image Recognition based on a LAB Feature Extraction Technique. Bul Inst Politehnic Iasi, Automat Control Comput Sci. 2011;LVII(LXI):33–41.

31 Porebski A, Vandenbroucke N, Macaire L. Haralick feature extraction from LBP images for color texture classification. 2008 IPTA. 2008:1–8. doi:10.1109/IPTA.2008.4743780

32 Kaur N, Nazir N, Manik. A Review of Local Binary Pattern Based texture feature extraction. 2021 ICRITO. 2021:1–4. doi:10.1109/ICRITO51393.2021.9596485

33 Figueiredo J, Medeiros G, Paula I Jr. Contour-based feature extraction for image classification and retrieval. 2016 SCCC. 2016:1–7. doi:10.1109/SCCC.2016.7836058

34 Koonce B. ResNet 50. In: Convolutional Neural Networks with Swift for TensorFlow. 1st ed. Berkeley (CA): Apress; 2021. p. 63-72. doi:10.1007/978-1-4842-6168-2_6.

CURRICULUM VITAE

NAME	Phanuphong Sutthawas
EDUCATIONAL BACKGROUND	
2021	Bachelor of Science Software Engineering Mae Fah Luang University