Pertanika J. Sci. & Technol. 33 (1): 283 - 312 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Machine Learning Precision for Mangosteen Maturity: A Comparative Analysis of Conventional Classifiers

Ram Kumar Thirumugam[1], Yasmin Mohd Yacob[1,2]*, Wan Mahfuzah Wan Ibrahim[3], Salina Mohd Asi[1,2] and Suhizaz Sudin[1]

[1]*Faculty of Electronic Engineering &Technology, Universiti Malaysia Perlis (UniMAP), Pauh Putra Campus, Arau, 02600, Perlis, Malaysia*
[2]*Centre of Excellence for Advanced Computing, Universiti Malaysia Perlis (UniMAP), Pauh Putra Campus, Arau, 02600, Perlis, Malaysia*
[3]*Horticulture Research Centre, Malaysian Agriculture Research Development Institute (MARDI), Sintok, Bukit Kayu Hitam, 06050, Kedah, Malaysia*

ABSTRACT

Identifying mangosteen maturity stages pre-harvest is crucial for postharvest quality, as fruit disease and pest infestation often occur at specific stages. Deep learning, while popular for classification, struggles with false negatives. Conversely, conventional machine learning methods now effectively handle false negative issues. The main goal of this research is to determine the significant comparison between different conventional classifiers, namely Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM) and K-Nearest Neighbour (K-NN), in terms of their accuracy, validity, and False Negative Rate (FNR) in predicting six distinct classes. Image samples of 253 mangosteens across six maturity stages were used, with 20 regions of interest (ROIs) each. 112 Gray-level Co-Occurrence Matrix (GLCM) and colour features were extracted to train models using texture, colour, and combined features. The evaluation metrics used for assessing the validity of predictions included precision, recall, F1-score, accuracy, and Cohen's Kappa. The RF classifier achieved high validation scores, with an accuracy of 0.76 and Cohen's Kappa of 0.70 for combined features, 0.75 and 0.69 for coloured features, and 0.46 and 0.33 for texture features. The Friedman test on false positive rates (FNR) across the four models shows significant differences ($p < 0.05$) for colour, texture, and their combination, with p-values of 0.00134, 0.00572, and 0.00071, respectively. RF is the best method, with the lowest mean FNR scores: 1.16 for texture, 1.16

ARTICLE INFO

Article history:
Received: 6 June 2024
Accepted: 3 September 2024
Published: 27 January 2025

DOI: https://doi.org/10.47836/pjst.33.1.13

E-mail addresses:
ramkumar@studentmail.unimap.edu.my (Ram Kumar Thirumugam)
yasmin.yacob@unimap.edu.my; yasminmyacob@gmail.com (Yasmin Mohd Yacob)
mahfuzah@mardi.gov.my (Wan Mahfuzah Wan Ibrahim)
salina@unimap.edu.my (Salina Mohd Asi)
suhizaz@unimap.edu.my (Suhizaz Sudin)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

for colour, and 1.00 for combined features. In conclusion, the RF classifier outperforms other classifiers in accuracy, validity, and mean FNR across six classes with three category features, achieving statistical significance in the Friedman Test.

Keywords: Conventional machine learners, decision tree, fruit maturity, K-Nearest Neighbour, random forest, support vector machine

INTRODUCTION

Mangosteen, technically known as Garcinia mangostana, is a tropical fruit native to Southeast Asia, notably Malaysia, Indonesia, and Thailand. Mangosteen is grown extensively in Malaysia's Pahang, Johor, and Perak states. Malaysia is a significant supplier of mangosteen worldwide, and the demand for this tropical fruit is growing, particularly in the United States, Europe, Japan, the Arab region, and ASEAN countries. Mangosteen is one of the nine premium fruits targeted by the Malaysian government, and mangosteen exports have increased by 65% since 2020 (IndexBox, 2024). The conventional method of assessing and identifying fruit quality through manual inspection by trained professionals is time-consuming and impractical for efficiently categorising large volumes of fruits intended for export within a limited time frame. To optimise the efficiency of fruit sorting and elevate the quality of categorisation and filtering, leveraging image processing and machine learning (ML) algorithms proves advantageous, as highlighted in studies by Pandey et al. (2013) and Liakos et al. (2018).

Recent trends employ deep learning (DL) in many aspects, including prediction and classification research areas, including crop grading, maturity or ripeness classification, and crop-yielding prediction (Khan, Nauman et al., 2022, Khan, Faheem et al., 2022; Joshua et al., 2022; Bashir et al., 2023; Orchi et al., 2023). Other studies which utilised DL in classifying leaf disease by Parashar and Johri (2024) and fruit maturity and defective fruits are by Al-Mashhadani and Chandrasekaran (2021), Mohtar et al. (2019), Kim et al. (2023), Ashtiani et al. (2021), Benmouna et al. (2022), Gao et al. (2020), Tan et al. (2020), Muñoz et al. (2022), Azizah et al. (2017) and Sudana et al. (2020). Nevertheless, according to Rushing (2022), the No Free Lunch (NFL) theorem claims that no particular ML algorithm is always the most effective for a given situation. Regardless of the contrasting notion among researchers, this work tends to agree that research utilising traditional machine learners should be continued so that improved techniques can be discovered in conventional machine learners to be, at the minimum, on par with deep learners' performance. Orchi et al. (2023) revealed that, despite the superior performance of the InceptionV3 network in terms of accuracy, precision, and recall, the Random Forest (RF) classifier demonstrated a markedly lower False Negative (FN) rate compared to other experimented models. Their findings are crucial for achieving heightened accuracy and mitigating misclassification rates.

In combining texture and coloured features, Phothisonothai and Tantisatirapong (2019) achieved 88% accuracy for mangosteen maturity stages within 3 classes. Whidhiasih et al. (2012) reached 85% accuracy across 6 maturity stages. Riyadi et al. (2020) achieved 94.16% accuracy, focusing on skin defects, not maturity stages. In addition, Parashar and Johri (2024) resulted in 94.76% accuracy in determining 4 classes of apple leaf disease. Several studies advocate for the incorporation of coloured features in fruit grading, as proposed by Khojastehnazhand et al. (2010) and Liming and Yanchao (2010). Damarjati et al. (2017) aced 80.8% accuracy using statistical features from mangosteen, but it focused on skin defects. Indeed, Afandi et al. (2021) scored 96.67% accuracy using texture features extracted via the Gray-level Co-Occurrence Matrix (GLCM) and executed using Extreme Learning Machines (ELM) for mangosteen. Nevertheless, the work detects skin defects. Another study by Riyadi et al. (2018) attained an accuracy of 92.5% utilising Discrete Cosine Transform (DCT) extracted texture features. However, the work detected mangosteen binary classes for skin defects unrelated to the fruit's maturity stages.

In a different category where numerical data were used, Khan, Faheem et al. (2022) achieved an accuracy of 96% and Joshua et al. (2022) with 97%, respectively. Khan, Nauman et al. (2022) and Bashir et al. (2023) gained accuracy of 78% and 91.44%, respectively, also using numerical data. These studies attempted to match the shape of the standard irrigation model.

The primary goal of this research is to compare the significance of various conventional classifiers, namely RF, DT, SVM and kNN, in terms of their accuracy and validity across various metrics such as accuracy, precision, recall, F1-score, and Cohen's Kappa when predicting six distinct classes. Next, this study aimed to analyse the processing power of the traditional machine learning models in relation to the false negative rate. In this area, deep learning models may fall short. This work attempts to understand the inherent strengths and limitations intrinsic to the classification methods implemented throughout this study, thus contributing to the advancement of knowledge not only of mangosteen maturity classification but also of other agricultural studies.

MANGOSTEEN CHARACTERISTICS

Mangosteen maturity is defined by various parameters such as size, colour, texture, and flavour, traversing through six discernible phases, each with unique characteristics. The first phase, Index Class 1, is marked by an immature yellowish-green colour, small size, and high mucus content. In Index Class 2, the fruit is half-developed green, continuing to mature with reduced mucus. Moving on to Index Class 3, the fruit's colour transitions to mature red-brownish, and the flesh becomes separable. Index Class 4 signifies ripeness, with the fruit maturing into red-purple hues, and the flesh is ready for consumption. Fully ripe in Index Class 5, the reddish mangosteen offers the finest flavour for immediate consumption. Finally, Index Class 6 indicates an overripe state, characterised by a purple-black colour, with mushy flesh. A visual representation of this multiclass progression of mangosteen maturity stages is depicted in Figure 1, as documented by Acharya et al. (2018) and Thammastitkul and Klayjumlang (2021), providing valuable insights into the intricacies of mangosteen development.

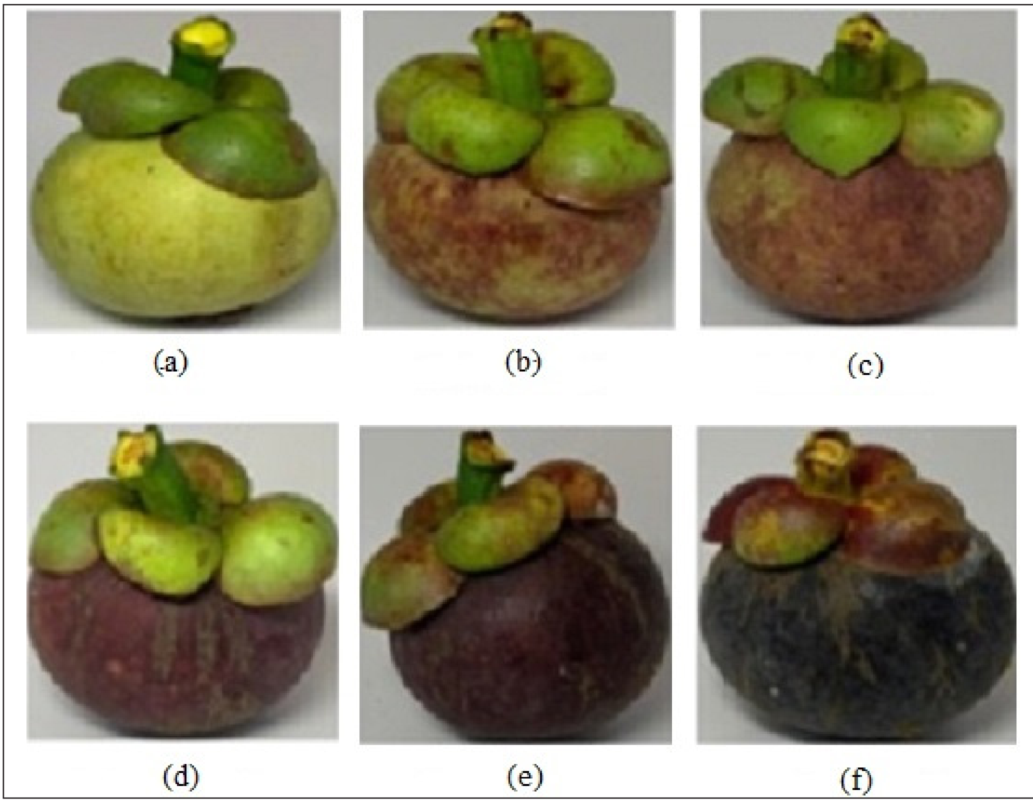

Figure 1. Mangosteen Maturity (a) Index Class 1; (b) Index Class 2; (c) Index Class 3; (d) Index Class 4; (e) Index Class 5; (f) Index Class 6

MACHINE LEARNING TECHNIQUES

The K-NN method, a supervised learning technique, evaluates new datasets by measuring their proximity to the k-nearest data points within the training set. It involves storing the training data and calculating the distance between each new and existing data point. Then, the K-NN algorithm identifies the k closest data points and assigns the new data point to the class containing most of these k nearest neighbours. Widely used for classification and regression tasks (Riyadi et al., 2020; Kim et al., 2023), the algorithm determines the projected label for a test sample in classification by selecting the label that predominates among the target labels of the k-selected training samples, as illustrated in Figure 2. K-NN's simplicity and effectiveness make it a popular choice in various fields, including pattern recognition, image recognition, and recommendation systems.

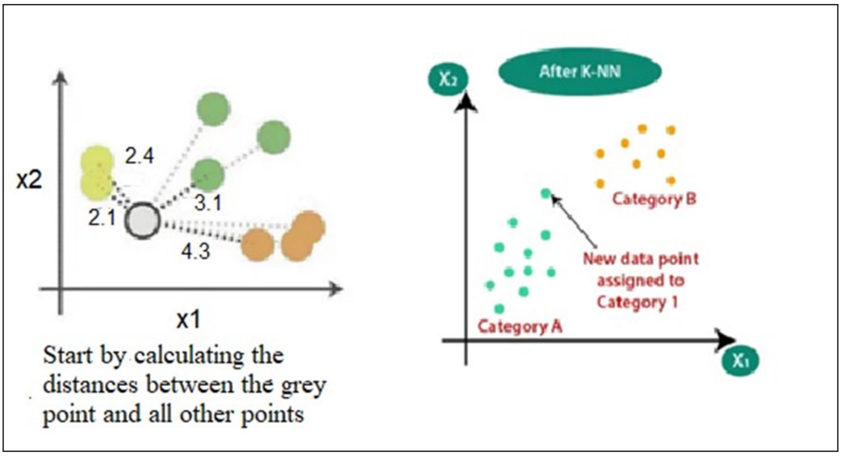

Figure 2. How K-NN determines classes (a) calculating distance (b) new data is classified

A Decision Tree (DT) serves as a tree-based methodology for decision-making and predictive modelling. It recursively partitions the data into subsets based on the most influential attribute and its corresponding value. This process continues until a decision or leaf node that denotes a prediction is reached. Each division in the tree represents a decision, and the path from the root to a leaf node delineates a sequence of decisions leading to a prediction. DTs exhibit versatility in handling categorical and numerical data, making them applicable to classification and regression tasks. Employing a hierarchical tree structure, the DT algorithm characterises datasets and computes discrete target-valued functions, as illustrated in Figure 3. The classification is accomplished by organising tree instances from the root to a leaf node. Meanwhile, each node of the tree corresponds to an attribute, and each branch signifies the value of that attribute. DTs stand as a robust tool widely employed across diverse domains for their interpretability and effectiveness in decision-making processes.

The Support Vector Machine (SVM) is a powerful supervised learning technique applicable to classification and regression endeavours. Its fundamental aim revolves around identifying the hyperplane that effectively segregates the data into separate classes or predicts target values accurately. This hyperplane's placement hinges on maximising the margin between it and the nearest data points, commonly known as support vectors. These support vectors, known as support vectors, serve as pivotal elements in delineating the hyperplane. These support vectors serve as pivotal elements in delineating the hyperplane and exert significant influence in making predictions for novel data points. The intricacies of SVM's operation and reliance on support vectors are vividly portrayed in Figure 4,

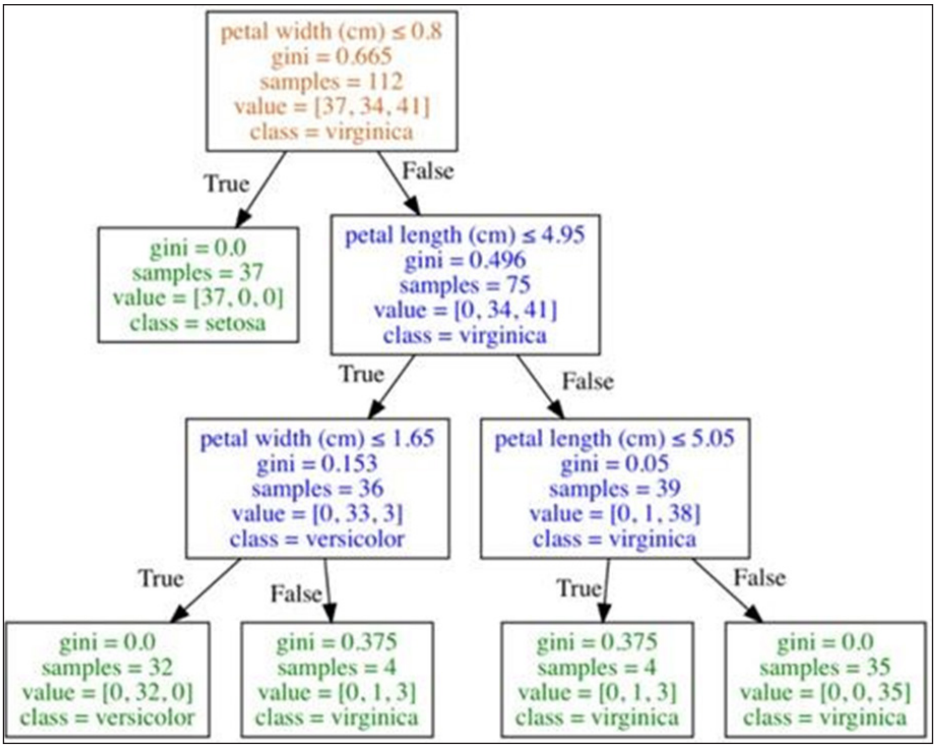

Figure 3. Construction of Decision Trees

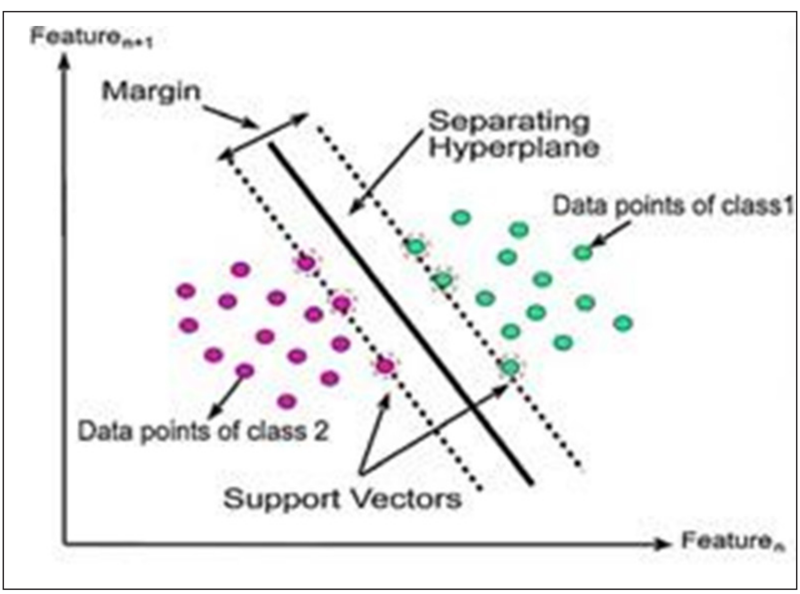

Figure 4. How a class is determined in Support Vector Machine (SVM)

elucidating its practical implementation and theoretical underpinnings (Kim et al., 2023). Notably, SVM's versatility extends beyond linear separation, encompassing non-linear relationships through techniques like kernel methods.

It is a cornerstone in various domains, including image recognition, bioinformatics, and financial forecasting, underscoring its indispensable role in modern machine learning paradigms.

In contrast, Random Forest (RF) emerges as an ensemble learning framework meticulously crafted for classification and regression tasks. Unlike singular decision trees, RF constructs numerous Decision Trees (DTs), amalgamating their predictions to yield a comprehensive outcome. Each tree within the forest stands as a unique entity, fashioned from a random subset of the dataset and a random selection of features. This diversity among the trees fortifies the model's resilience and predictive prowess, mitigating the risk of overfitting. The ultimate prediction is forged by amalgamating outcomes from all constituent trees in the random forest, elucidated in Figure 5 (Goel & Abhilasha, 2017; Ashtiani et al., 2021). RF's versatility extends beyond linear separations, facilitating the discernment of complex, non-linear relationships present in data. Its efficacy and adaptability render it a cornerstone in modern machine learning, finding applications in domains ranging from finance to healthcare, underpinning its status as a quintessential tool in predictive analytics.

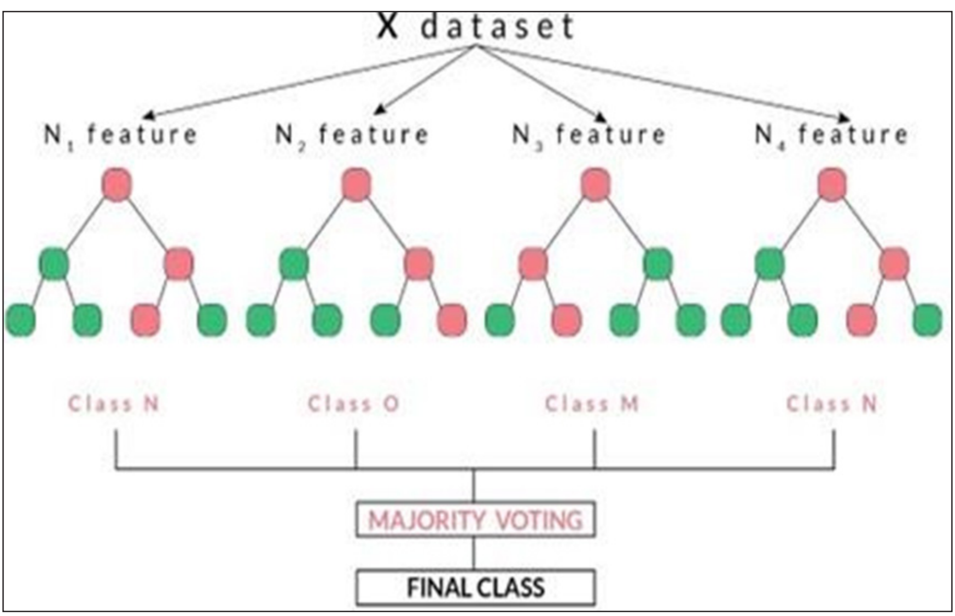

Figure 5. Concept of Random Forest

METHODOLOGY

Image Acquisition

The mangosteen samples for this study were provided by the Malaysian Agriculture Research Development Institute (MARDI) in Changlun, Kedah, Malaysia. Postharvest, these specimens underwent photographic documentation at the Post Harvest Department, as seen in Figure 6a and Figure 6b. Image acquisition employed a Do-It-Yourself (DIY) mini-studio setup, outlined in Figure 6c, capturing the external surface of the mangosteen. A light source ensured optimal visibility. Despite a limited number of images, 20 regions of interest (ROIs) were extracted from each, standardised to 512 x 512 pixels. Table 1 outlines the distribution of images and ROIs across different maturity classes: Class 1 contributed 25 images and 525 ROIs, Class 2 had 14 images and 294 ROIs, Class 3 contained 48 images and 1008 ROIs, while Class 4 comprised 62 images and 1302 ROIs. Classes 5 and 6 included 51 and 53 images, contributing 1071 and 1113 ROIs, respectively. From each image, textured and coloured features were extracted, resulting in 5313 ROIs or instances.

This study embraced a hybrid approach to feature extraction, drawing from methodologies proposed by Afandi et al. (2021) for textural features and leveraging coloured feature extraction techniques outlined in several studies (Khojatesnazhand et al., 2010; Leemans et al., 2002; Liming & Yanchao, 2010). Textural feature extraction commenced with converting each image into grayscale and then extracting 20 Regions of Interest (ROIs) from each rendition. Conversely, for coloured feature extraction, a comprehensive strategy unfolded, tapping into four distinct colour models: RGB, hue, saturation, value (HSV), CIE Lab, and CIE LUV. Each image transformed its designated colour model, with 20 ROIs extracted from each colour channel within every model. This dual-pronged approach facilitated an exhaustive examination of textured and coloured attributes, enriching subsequent analysis. By amalgamating these techniques, the study aimed to provide a nuanced understanding of the intricate interplay between texture and colour characteristics in the dataset under scrutiny.

(a) (b) (c)

Figure 6. Activities of image acquisition

Table 1
Number of images and ROIs for each maturity class

Class	Images	ROI
1	25	525
2	14	294
3	48	1008
4	62	1302
5	51	1071
6	53	1113

Extracted Features from Mangosteen Images

Textured features were meticulously extracted using the Gray-Level Co-occurrence Matrix (GLCM) method pioneered by Haralick et al. (1973), encompassing 13 distinctive features. These included Energy, Contrast, Correlation, Variance, Homogeneity, Sum average, Sum variance, Sum entropy, Entropy, Difference Variance, Difference Entropy, Measure of Correlation 1, and Measure of Correlation 2. Each textured attribute was derived from four directional angles: 0, 45, 90, and 135 degrees, resulting in 52 textured features per Region of Interest (ROI). Conversely, coloured features encompassed mean, standard deviation, variance, entropy, and Root Mean Square (RMS) extracted from each ROI across every colour channel and colour model. Consequently, 60 coloured features were obtained, summing up to 112 features from textured and coloured domains for every ROI. This comprehensive feature extraction approach aimed to capture diverse characteristics embedded within the dataset, laying the groundwork for in-depth analysis and exploration of intricate patterns and correlations.

Experimentation Setup

The experimentation process encompassed three distinct feature modes: textured features only, coloured features only, and a fusion of both textured and coloured features. This comprehensive approach aimed to dissect the significance of each feature type within the designated classifiers, facilitating a better understanding of their respective contributions. Table 1 delineated a substantial dataset comprising 5313 Regions of Interest (ROIs) or instances amassed for this study. The experimental protocol comprised a bifurcated approach, with 70% of the instances allocated for training purposes, while the remaining 30% constituted the test set. Instances were meticulously selected across various maturity classes to ensure a representative distribution within training and testing subsets. Consequently, 3,719 mangosteen ROI images spanning maturity class 1 through maturity class 6 were earmarked for training, supplemented by an additional 1,594 mangosteen ROI images earmarked for testing. Figures 7a to 7f showcase sample ROIs representing mangosteen maturity classes 1 through 6 in grayscale. Figures 7g to 7l present analogous

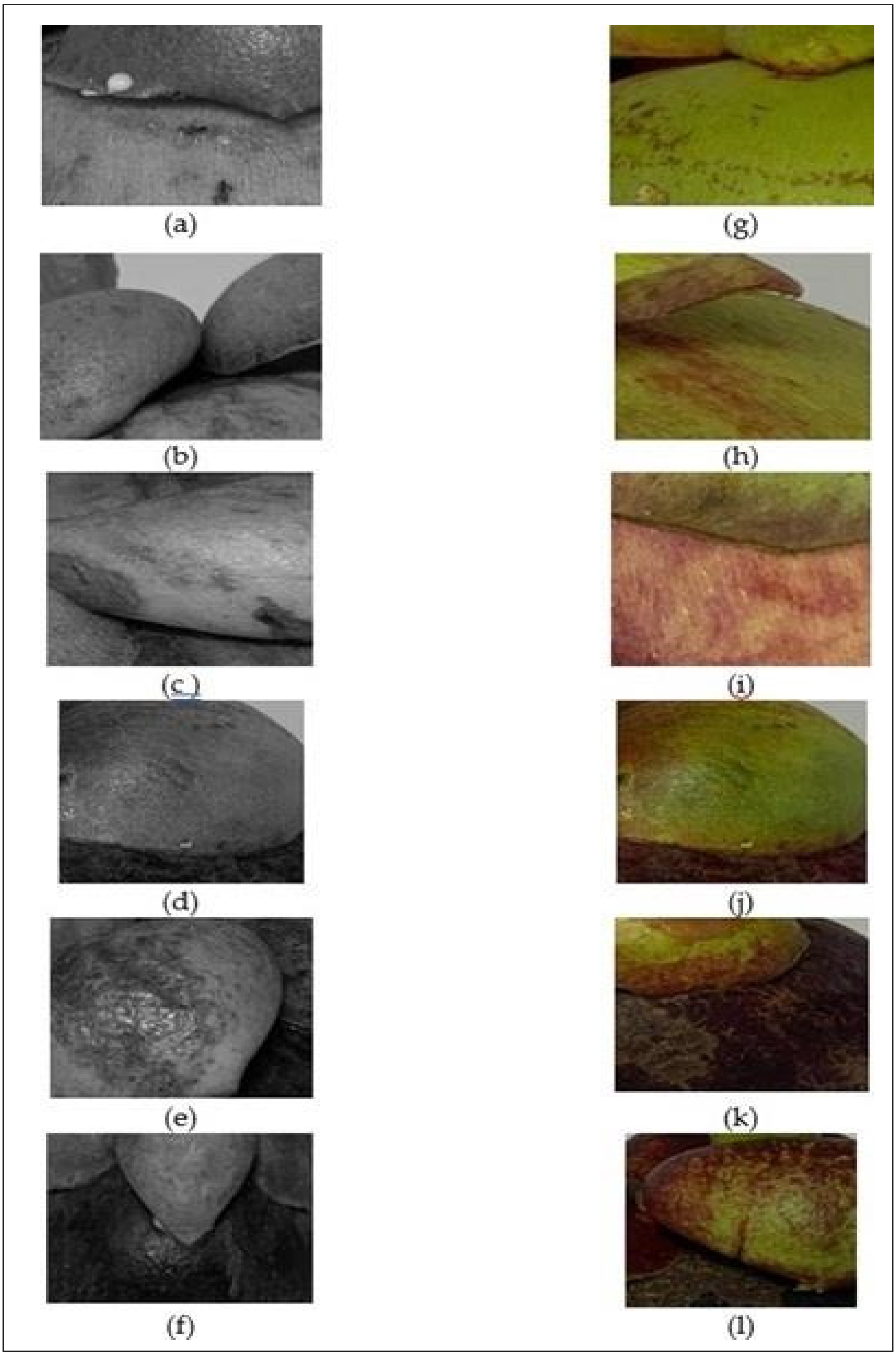

Figure 7. Samples of ROIs for each maturity class 1-6 respectively (a-f) in grayscale and (g-l) in RGB colour model

maturity stages in the RGB colour model, offering insights into the visual characteristics across different maturity levels.

The comprehensive processes delineated in Figure 8 encompass various stages, commencing with image acquisition, which involves the meticulous capture and subsequent conversion of images into grayscale and coloured formats. Following this, images undergo precise cropping to predefined dimensions, yielding Regions of Interest (ROIs) as an integral part of the image pre-processing pipeline, paving the way for subsequent analyses. Feature extraction ensues, wherein a rich array of both textured and coloured features is meticulously acquired and catalogued, subsequently categorised into three distinct datasets.

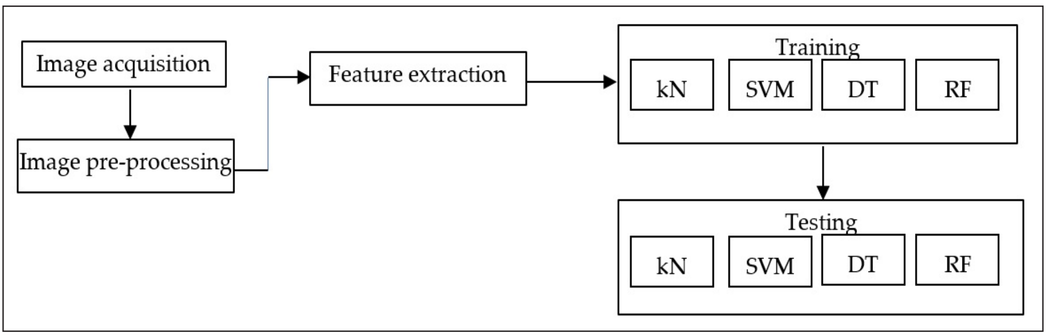

Figure 8. The flow of processes in the experiment

Modes: texture-only, colour-only, and a hybrid amalgamation of textured and coloured features. Each dataset undergoes rigorous training utilising four distinct classifiers: K-Nearest Neighbours (K-NN), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). Remarkably, each classifier is adeptly leveraged across all three dataset modes, signifying the versatility and adaptability of the approach. This meticulous selection of classifiers is underpinned by their demonstrated efficacy in addressing multiclass problems, as underscored by previous research endeavours (Giuntini et al., 2023; Boateng et al., 2020), highlighting their suitability for the present study's objectives.

In this study, the specified classifiers accessible within the Matlab tool were employed for analysis. Furthermore, each type of the remaining dataset underwent thorough testing utilising the K-Nearest Neighbours (K-NN), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) classifiers, respectively. Each machine learner performs the experimentation based on the parameter setting described in Table 2. The research findings are meticulously presented based on a comprehensive analysis of a testing dataset, employing diverse evaluation metrics to ensure a robust assessment. These chosen evaluation metrics encompass a broad spectrum, including accuracy, precision, recall, F1-score, Cohen's Kappa, and false negative rate (FNR), allowing for a multifaceted classifier performance evaluation. Particularly in the context of FNR analysis, a detailed examination was conducted by constructing confusion matrices for each feature mode. Within these matrices, the counts of false negatives and true positives were meticulously tallied for each maturity class across all the aforementioned classifiers, providing invaluable insights into the classification outcomes and performance metrics.

Table 2
Parameter setting for experimented machine learners

	N_estimators	**100**	**Decision Tree**	**Criterion**	**Gini**
Random Forest	Criterion	Gini		Splitter	Best
	Random state	None			
	Kernel	RBF		Number of neighbours	Default=5
SVM	Probability	False	kNN	Weight	Uniform
	Random state	None		Leaf size	Default=30
	Stratify	Yes		Metric	Default = 'Minkowski'

Evaluation Metrics

In addition to accuracy, as defined in Equation 1, this study incorporated diverse evaluation metrics. Following the recommendations of Powers (2020), precision, recall, F1-score, and Cohen's Kappa were chosen as suitable metrics for assessing classifier performance in multiclass scenarios. Equation 1 quantifies accuracy, representing the ratio of true cases to overall cases. Precision, detailed in Equation 2, measures the quality of results by evaluating true positives against the sum of true positives and false positives. Equation 3 defines recall as the ratio of true positives to the sum of true positives and false negatives, with precision gauging quality and recall gauging quantity. A higher precision indicates a model returning more relevant results than irrelevant ones, while a higher recall signifies a model retrieving the most relevant results irrespective of irrelevant ones. The F1-score, depicted in Equation 4, serves as the harmonic mean of precision and recall, providing a balanced evaluation of both metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad [1]$$

$$Precision = \frac{TP}{TP + FP} \quad [2]$$

$$Recall = \frac{TP}{TP + FN} \quad [3]$$

$$F1\ score = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad [4]$$

The F1 score also represents the model's accuracy on a dataset. F1-score ranges between 0 and 1. The closer it is to 1, the better the model. Cohen's Kappa measures the level of agreement between two raters who each classify items into mutually exclusive categories. In other words, the measure tests interrater reliability. The rater reliability represents the extent to which the data collected in the study are correct representations of the variables measured. Cohen's Kappa, namely k, is described in Equation 5. Cohen's Kappa always ranges between zero and one, with zero indicating no agreement between the two raters and one indicating perfect agreement between the two raters.

$$k = \frac{P_o - P_e}{1 - P_e} \qquad [5]$$

where P_o: relative observed agreement among raters

P_e: hypothetical probability of chance agreement

The False Negative Rate (FNR) denotes instances where a true condition is inaccurately identified as false. For example, it quantifies how often instances are wrongly predicted as non-class 1 when they indeed belong to class 1 in the dataset. The FNR is calculated using Equation 6, whereby a smaller FNR is a better value given that false negative (FN) and true positive (TP) cases have been determined from the confusion matrix. A true positive signifies the model's accurate prediction of a specific class.

$$FNR = \frac{FN}{FN + TP} \qquad [6]$$

According to Rainio et al. (2024), a suitable statistical test for multiclass classification that determines significant differences among several models is the Friedman Test. Given that X^2 is chi-square, N is a number of instances, df is the degree of freedom, and the p-value is the area under the density curve of the chi-square distribution to the right of the value of the test statistic. If the p-value is less than the significance level, there is enough evidence to conclude that at least one mean of the chosen metric is different among the multiple methods. This work chooses 0.05 as the significance level because it is common in classification. The lowest mean ranks signify the best statistically significant method. It is calculated by taking the sum of the ranks and divided by N.

RESULTS

Results Related to Accuracy

The study found that most learners achieved less than 50% accuracy, as evidenced by the results from the texture features testing data in Table 3. Random Forest (RF) and Decision Tree (DT) precisely detected class 1 maturity within the range of 0.58 to 0.61. Support

Table 3
Results analysed from testing data using various metrics from texture, colour and combined texture-colour features

Classifier	Class	Precision			Recall			F1-Score			Accuracy			Cohen's Kappa		
		Texture	**Colour**	**Combined**	**Texture**	**Colour**	**Combined**	**Texture**	**Colour**	**Combined**	**Texture**	**Colour**	**Combine**	**Texture**	**Colour**	**Combined**
Random Forest	1	0.58	0.78	0.76	0.58	0.9	0.9	0.58	0.84	0.82	0.46	0.75	0.76	0.33	0.69	0.70
	2	0.41	0.64	0.68	0.22	0.5	0.45	0.29	0.56	0.55						
	3	0.55	0.8	0.78	0.61	0.76	0.8	0.58	0.78	0.79						
	4	0.45	0.65	0.68	0.47	0.66	0.67	0.46	0.66	0.67						
	5	0.34	0.65	0.67	0.35	0.67	0.69	0.35	0.66	0.68						
	6	0.44	0.93	0.94	0.42	0.91	0.9	0.43	0.92	0.92						
Decision Tree	1	0.61	0.84	0.74	0.58	0.77	0.78	0.59	0.8	0.76	0.4	0.69	0.67	0.26	0.62	0.59
	2	0.29	0.49	0.42	0.29	0.58	0.47	0.29	0.53	0.44						
	3	0.47	0.71	0.66	0.47	0.68	0.63	0.47	0.69	0.65						
	4	0.4	0.59	0.56	0.39	0.59	0.54	0.39	0.59	0.55						
	5	0.33	0.61	0.61	0.36	0.62	0.62	0.34	0.62	0.62						
	6	0.36	0.87	0.88	0.35	0.9	0.88	0.35	0.88	0.88						
SVM	1	0	0.71	1	0	0.77	0.01	0	0.74	0.01	0.26	0.66	0.26	0.03	0.57	0.04
	2	0	0	0	0	0	0	0	0	0						
	3	0.33	0.65	0.4	0.01	0.74	0.01	0.01	0.69	0.01						
	4	0.25	0.56	0.25	0.87	0.75	0.86	0.39	0.64	0.39						
	5	0.32	0.63	0.3	0.21	0.3	0.22	0.26	0.41	0.25						
	6	0	0.78	0.7	0	0.92	0.04	0	0.84	0.08						
K-NN	1	0.18	0.6	0.32	0.17	0.76	0.31	0.17	0.67	0.32	0.28	0.64	0.36	0.09	0.55	0.19
	2	0.1	0.41	0.22	0.07	0.36	0.17	0.08	0.38	0.19						
	3	0.33	0.65	0.41	0.33	0.65	0.42	0.33	0.65	0.42						
	4	0.28	0.59	0.32	0.39	0.56	0.44	0.33	0.57	0.38						
	5	0.3	0.57	0.31	0.3	0.54	0.31	0.3	0.56	0.31						
	6	0.26	0.8	0.48	0.18	0.83	0.3	0.21	0.81	0.37						

Vector Machine (SVM) excelled in class 4 detection, achieving a recall value of 0.87. RF and DT performed well in detecting stages other than class 2, with recalls ranging from 0.58 to 0.22, which is understandable due to SVM's binary class inclination. Considering accuracy and Cohen's Kappa, RF outperformed other classifiers.

Moving forward, Table 3 offers a comprehensive overview of the results garnered from testing data, utilising an array of metrics derived from colour features. Random Forest (RF) exhibited exceptional performance, boasting an accuracy of 0.75 and a Cohen's Kappa value of 0.69. RF showcased its superior ability in pinpointing both class 1 and class 6, achieving recall values that surpassed 0.90. Intriguingly, nearly all classifiers demonstrated proficiency in determining class 6 using colour features, highlighting the significance of colours within this category. Furthermore, RF displayed precision in identifying each class, with values ranging from 0.64 to 0.93. However, it is noteworthy that Decision Tree (DT) outperformed RF in detecting class 1, achieving a precision value of 0.84 compared to RF's 0.78. Nevertheless, the overall performance of RF remains justified, given its voting approach to determine specified classes.

Table 3 provides a detailed overview of the results from analysing testing data, utilising a blend of texture and colour features. The findings prominently showcase the Random Forest (RF) model's superiority, boasting an impressive accuracy of 0.76 and a Cohen's Kappa value of 0.70. RF's outperformance across F1 scores for each class is noteworthy, indicating its effectiveness in distinguishing various mangosteen maturity classes through the harmonic mean of combined precision and recall. Conversely, the Support Vector Machine (SVM) falls short, with a lower accuracy rate of 0.26 and a Cohen's Kappa value of 0.04. SVM notably struggles in accurately identifying maturity class 2.

Statistical Test Result Based on F1-score

Table 4 presents the Friedman Test result to discover the p-value and mean rank for 3 categories of features across 6 classes of maturity stages based on F1-score metrics. Instead of accuracy, the F1-score was chosen to perform the statistical test because it represents a balanced evaluation of precision and recall. The p-value for texture features is 0.00235, which is less than 0.05 level of significance. Thus, there is enough evidence to conclude that at least one mean F1-score is different among the multiple classifiers under the texture features category. The mean rank for texture features using classifiers RF, DT, SVM, and kNN are 3.75, 3.16, 1.25, and 1.83, respectively. On the other hand, the p-value for colour features is 0.00235, which is less than 0.05 level of significance. Thus, there is enough evidence to conclude that at least one mean F1-score is different among the multiple classifiers under the colour features category. The mean rank for colour features using classifiers RF, DT, SVM, and kNN are 4.00, 2.75, 1.91, and 1.33, respectively. The p-value for a combination of texture and colour features is 0.00071, which is less than 0.05 level of significance. Thus, there is enough evidence to conclude that at least one mean F1-score is different among the multiple classifiers under the combined features category.

The mean rank for combined features using classifiers RF, DT, SVM, and kNN are 4.00, 3.00, 1.16, and 1.83, respectively.

Table 4
Friedman Test result based on F1-score related to texture, colour and combined features

	Classifier	Sum of Ranks	X2	df	N	p-value	Mean Rank
Texture	RF	22.50	14.45	3	6	0.00235	3.75
	DT	19.00					3.16
	SVM	7.50					1.25
	kNN	11.00					1.83
Colour	RF	24.00	14.45	3	6	0.00235	4.00
	DT	16.50					2.75
	SVM	11.50					1.91
	kNN	8.00					1.33
Combined colour and texture	RF	24.00	17.00	3	6	0.00071	4.00
	DT	18.00					3.00
	SVM	7.00					1.16
	kNN	11.00					1.83

Results Related to Confusion Matrix and FNR

Figure 9 presents the confusion matrix depicting the results of testing textured features using various classifiers. Figure 10 illustrates qualitative measures, specifically the FNR derived from the confusion matrix in Figure 9. Random Forest (RF) exhibits the lowest false negative rate among the six maturity classes. Support Vector Machine (SVM) shows proficiency in determining maturity stages for classes 3, 4, and 5, with false negative rates ranging between 0.6 and 0.8. However, SVM performs badly in other classes.

The overall misclassification rate for the textured features scheme generated by RF, DT, SVM and kNN was 53.8%, 59.6%, 74.2% and 72.5%, respectively. The overall accuracy for the coloured features scheme using RF, DT, SVM and kNN were 75%, 69.4%, 65.6% and 63.6%, respectively. The overall accuracy for the combined texture-colour features scheme using RF, DT, SVM and kNN were 76%, 66.9%, 26.5% and 35.5%. The gap based on FNR between RF-DT and SVM-kNN scheme based on textured features in Figure 10 was huge because of the higher misclassification rate between them.

Figure 11 presents the confusion matrix depicting the results of testing-coloured features using various classifiers. Figure 12 provides a visual representation of the FNR derived from the confusion matrix in Figure 11. Random Forest (RF) emerges as the standout performer within this visualisation, boasting the lowest FNR across all classes, with rates between 0.1 and 0.4. The Support Vector Machine (SVM) encountered difficulties accurately detecting the class 2 maturity stage. Undoubtedly, RF distinguishes itself as the most promising learner for discerning the various maturity classes of mangosteen within the dataset under scrutiny. This assertion finds support in its impressive accuracy, Cohen's Kappa scores, and its consistently low false negative rates. Although SVM and kNN missed the 0.60 acceptance value based on Cohen's Kappa, all the classifiers in the study achieved more than 0.60 accuracies. These findings suggested that colour features better describe mangosteen multiclass maturity. Gaps based on FNR between RF, DT, SVM and kNN based on coloured features in Figure 12 were reduced compared to the one in Figure 10, which employed textured features because of the lower misclassification rate among them.

Figure 13 presents the confusion matrix depicting the results of testing combined features using various classifiers. Figure 14 illustrates qualitative measures, specifically the FNR derived from the confusion matrix in Figure 13. Notably, the Random Forest (RF) exhibits the lowest FNR across classes, ranging from 0.05 to 0.32. This exceptional performance by RF, characterised by consistently low FNR values, complements the overall evaluation based on accuracy and Cohen's Kappa. In contrast, the Support Vector Machine (SVM) faces challenges in determining class 1 and class 2 maturity, which is evident in its higher FNR. Intriguingly, despite SVM having the lowest FNR across classes compared to kNN, it surprisingly outperforms kNN in determining class 6. The gap based on FNR between RF-DT and SVM-kNN scheme based on combined texture-colour features in Figure 14 was huge because of the higher misclassification rate between them.

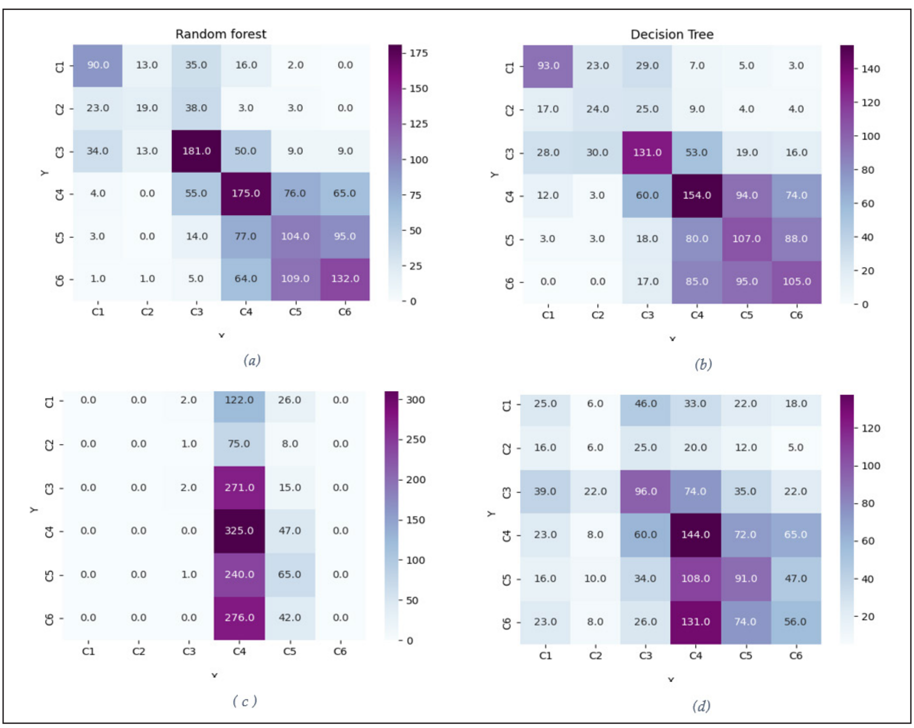

Figure 9. Confusion matrix with results analysed from the testing data of textured features executed via various classifiers

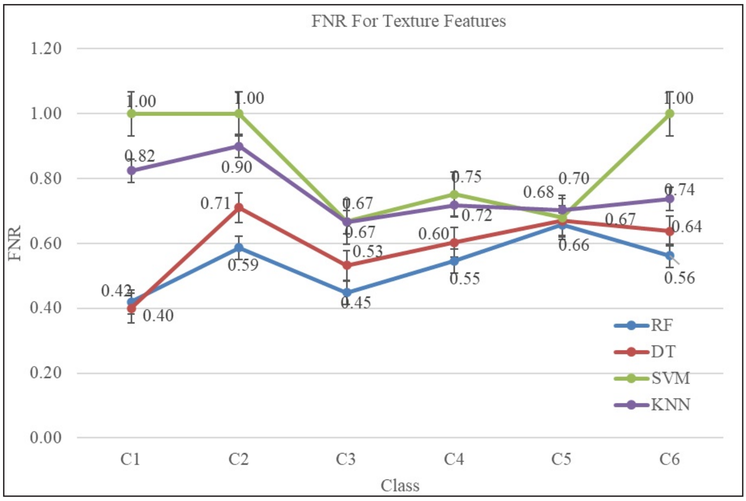

Figure 10. False negative rates analysed from the testing data of textured features executed via various classifiers

Figure 11. Confusion matrix with results analysed from the testing data of coloured features executed via various classifiers

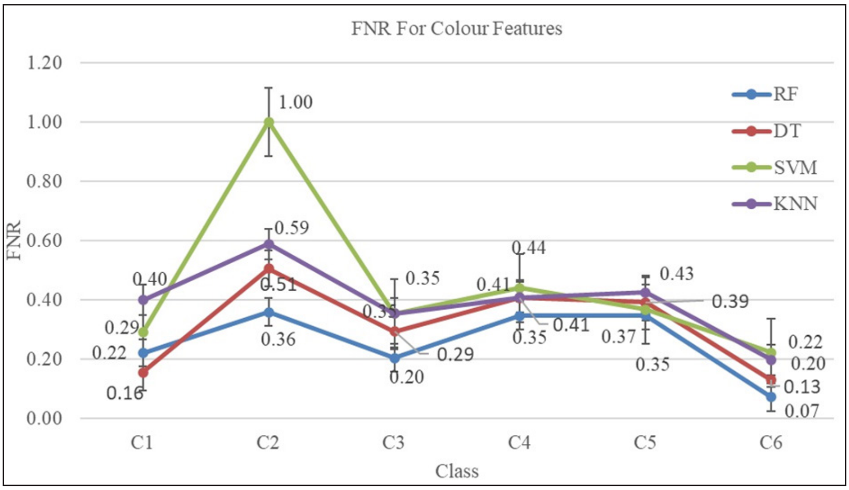

Figure 12. False negative rates from the testing data of coloured features executed via various classifiers

Figure 13. Confusion matrix with results analysed from the testing data of combined textured and coloured features executed via various classifiers

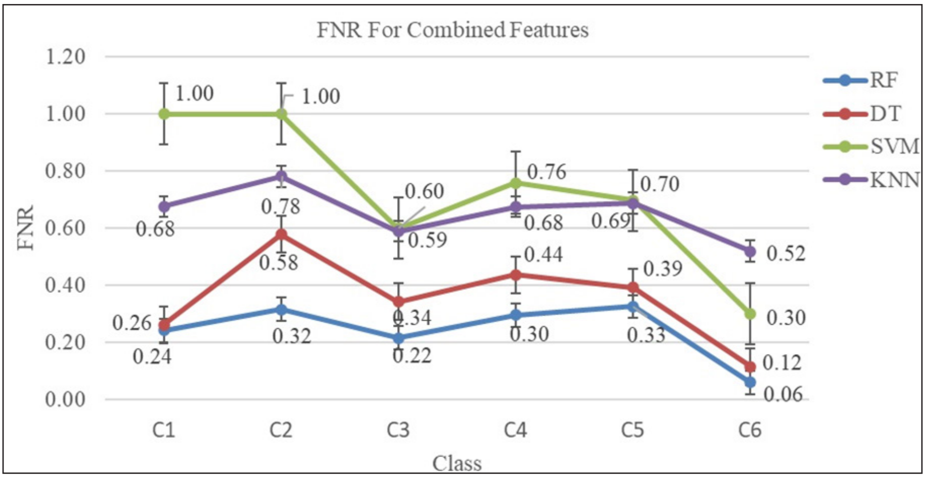

Figure 14. False negative rate from the testing data of combined textured and coloured features executed via various classifiers

Statistical Test Result Based on FNR

Table 5 presents the Friedman Test result to discover the p-value and mean rank for 3 categories of features across 6 classes of maturity stages based on FNR metrics. The p-value for texture features is 0.00134, which is less than 0.05 level of significance. Thus, there is enough evidence to conclude that at least one mean FNR is different among the multiple classifiers under the texture features category. The mean rank for texture features using classifiers RF, DT, SVM, and kNN are 1.16, 1.83, 3.75, and 3.25, respectively. On

Table 5
Friedman Test result based on FNR related to texture, colour and combined features

	Classifier	Sum of Ranks	X2	df	N	p-value	Mean Rank
Texture	RF	7.00	15.65	3	6	0.00134	1.16
	DT	11.00					1.83
	SVM	22.50					3.75
	kNN	19.50					3.25
Colour	RF	7.00	12.55	3	6	0.00572	1.16
	DT	12.50					2.08
	SVM	20.50					3.42
	kNN	20.00					3.33
Combined colour and texture	RF	6.00	17.00	3	6	0.00071	1.00
	DT	12.00					2.00
	SVM	23.00					3.83
	kNN	19.00					3.17

the other hand, the p-value for colour features is 0.00572, which is less than 0.05 level of significance. Thus, there is enough evidence to conclude that at least one mean FNR is different among the multiple classifiers under the colour features category. The mean rank for colour features using classifiers RF, DT, SVM, and kNN are 1.16, 2.08, 3.42, and 3.33, respectively. The p-value for a combination of texture and colour features is 0.00071, which is less than 0.05 level of significance. Thus, there is enough evidence to conclude that at least one mean FNR is different among the multiple classifiers under the combined features category. The mean rank for combined features using classifiers RF, DT, SVM, and kNN are 1.00, 2.00, 3.83, and 3.7, respectively.

DISCUSSION

Discussion Related to Accuracy

The analysis of various metrics indicates a notable similarity in the accuracy of Random Forest (RF) and Decision Trees (DT) when utilising coloured features alone and in combination with textured features. The experimental findings strongly suggest that the use of colour features alone is sufficiently robust for characterising mangosteen maturity classes. Specifically, incorporating thirteen out of fourteen textural features from GLCM and five essential colour features—mean, standard deviation, variance, entropy, and RMS—from each colour channel and model (RGB, HSV, CIE Lab, and CIE LUV) significantly contributes to accurate maturity class determination for mangosteen. RF is a superior classifier to DT, SVM and kNN due to its less biased nature—RF's performance benefits from majority voting across multiple sub-tree iterations.

The results based on the F1-score among all the compared machine learners are statistically significant for each type of feature scheme, as the p-value for all feature schemes is less than the 0.05 significance level. However, RF is not the most statistically significant method based on the mean rank values in Table 6. This finding is consistent with the methods compared to other works that employed deep learning (DL), as shown in Table 6. For instance, Sudana et al.'s (2020) work determined mangosteen maturity for seven classes using coloured features and achieved 97.10% accuracy. Another study by Mohtar et al. (2019) achieved 91.90% accuracy using Inception V3 to determine six classes of mangosteen maturity via coloured features. Although the work by Parashar and Johri (2024) is not directly comparable because it determined four classes of apple leaf disease, the Inception V3 model achieved an accuracy of 94.76%.

Traditional machine learning methods for determining maturity, skin defect, or ripeness have not performed as well as most deep learning results. Referring to Table 6, for example, Phothisonothai and Tantisatirapong (2019) conducted a study to determine mangosteen ripeness for three classes using texture and coloured features via GMM, achieving 86.67% accuracy. Riyadi et al. (2020) studied mangosteen binary classes of skin defects using texture and coloured features, scoring 94.16% accuracy via LDA. Whidhiasih et al. (2012) achieved 85% accuracy in determining three classes of mangosteen maturity using textured and coloured features classified via Fuzzy NN.

As observed in Table 6, many studies did not report precision and recall, leaving insufficient information to conclude the FNR performance. Orchi et al.'s (2023) work compared DL and traditional learners, including Inception V3, RF, SVM, and K-NN, but did not determine FNR and focused on binary classes of crop leaf disease. In his work, RF achieved a competitive accuracy of 97.54% compared to Inception V3 accuracy of 98.01%.

Regardless of RF's ability to classify multiclass instances, the instances acquired from textured features did not really represent the overall characterisation of the mangosteen maturity stages. Besides texture, mangosteen maturity is also characterised by their skin colour. DT performed worse than RF but was much better than SVM and K-NN, using textured features only. It is supported by the fact that DT suffers a bias problem. However, because the classifier is suitable for multiclass problems, the DT recognition rate is much better than that of SVM and K-NN. It is observed that having textured features only is insufficient to characterise the mangosteen maturity classes.

It was mentioned earlier that the overall accuracy for the textured features scheme using RF was 46.2%. Meanwhile, when coloured features were employed, the recognition rate increased to 75% using the same classifier. This result indicates that coloured features are significantly better at characterising mangosteen maturity stages compared to textured features. A similar trend was observed with the remaining classifiers. The overall accuracy for the coloured features scheme using both SVM and K-NN ranged between 63% and 66% for the six mangosteen maturity classes. From earlier findings, RF and DT achieved an overall accuracy of 69% and 75% for the coloured features scheme. It suggests that coloured features are essential for characterising mangosteen maturity classes regardless of the classifier used, though some classifiers perform better or worse depending on their processing capabilities.

The combined coloured texture feature was 76% using the same type of classifier. Coloured features or a combination of colour texture features are suggested to characterise mangosteen maturity classes.

Surprisingly, the overall accuracy of the coloured features and combined texture-colour features scheme using the same DT classifier was 69.4% and 66.9%, respectively. Indeed, the use of coloured features and combined texture-colour features are adequate to characterise mangosteen maturity stages.

The overall accuracy for the colour features scheme and combination of texture-colour using SVM was 65.6% and 26%, respectively. Meanwhile, the overall accuracy for the colour features scheme and combination of texture-colour using K-NN was 63.6% and 35.5%, respectively. There is a huge decrease in the overall accuracy when combined texture-colour features are used and executed using SVM and K-NN. These findings may be because those classifiers have less ability to handle high dimensional data, thus less processing power.

Based on the experimentation result, this work suggests that using colour features is sufficient to characterise mangosteen maturity classes. Alternatively, combined texture-colour features are better executed using an RF classifier to determine mangosteen maturity classes. Indeed, five coloured features, namely mean, standard deviation, variance, entropy, and RMS from each colour channel and each colour model of RGB, HSV, CIE Lab, and CIE LUV, are essential to determine maturity classes for mangosteen. RF classifier can handle multiclass problems, which is a valuable notion in this kind of work. In addition, RF lessens the overfitting problem in DT and reduces the variance, improving accuracy. On top of that, non-linear parameters are handled efficiently, and RF is robust to outliers.

Discussion Related to FNR

The False Negative Rates (FNR) for RF, whether using coloured features alone or combined texture and coloured features, exhibit consistent similarity, ranging from 0.02 to 0.40. It suggests that determining mangosteen multiclass maturity stages can be effectively achieved using either coloured features or a combination of texture and colour features. From another perspective, combined features are not the factor degrading SVM and K-NN multiclass detection capability. The competitive accuracy values support it, as Cohen's Kappa and FNR were obtained from RF and DT. RF achieved an accuracy of 0.76 and DT of 0.67. Both RF and DT Cohen's Kappa readings are above 0.59, with RF having a higher reading of 0.70. The FNR for RF is more than 0.60 but is considered competitive when it falls between 0.02 and 0.40.

Due to their inherent limitations, SVM and K-NN are less effective in determining multiclass mangosteen maturity. SVM is more suited for binary classification, while K-NN, although supporting multiclass cases, struggles to differentiate them effectively using the proximity concept. K-NN does not require training before classification, whereas accurately differentiating among mangosteen maturity classes necessitates a learning process.

This observation aligns with the results presented in Table 3, where RF's performance is rooted in the majority voting concept. SVM is known to better suit binary classification. In this work, kNN, as a lazy learning algorithm, falls short in finding the similarity of new and existing points in specified classes.

Once again, RF's commendable performance can be attributed to its robust computing capability, which employs a voting mechanism among subtrees during iteration. This unique approach mitigates bias and proves highly effective in accurately classifying multiclass cases.

This anomaly can be attributed to SVM's nature as a binary classifier. It distinguishes between class 2 and non-class 2 categories, potentially contributing to its superior performance in certain instances. This nuanced understanding of classifier behaviour enhances our insights into their strengths and limitations in differentiating various maturity classes.

Regardless of the FNR rate, RF and DT classify better in all feature schemes. SVM and K-NN have less processing power to classify when only textured and a combination of texture-colour features are used. Regardless of the feature scheme, RF is always the best classifier compared to DT. Additionally, RF is the best at classifying maturity stages among all the classifiers.

Results based on FNR among all the compared machine learners in Table 5 are statistically significant for each type of feature scheme because the p-value for all feature schemes is less than the 0.05 significance level. RF is also the most statistically significant method, scoring the lowest mean rank values in Table 5.

The competitive performance of FNR using the RF classifier across multiple classes highlights its adeptness in handling multiclass problems, a crucial attribute in this context. Furthermore, RF effectively mitigates the overfitting issues associated with DT, reduces variance, and thereby enhances overall accuracy. Its proficiency in managing non-linear parameters and resilience to outliers make RF a robust choice for this analytical task.

When comparing the experimented methods with previous work on statistically significant methods based on FNR, there is insufficient information to conclude because many previous works did not compute precision and recall metrics.

CONCLUSION

RF is deemed the best method based on its accuracy in determining six mangosteen maturity classes compared to the experimented methods. A statistical test, the Friedman Test, was conducted using the F1-score, and it was found that at least one method is statistically significant and differs among the multiple methods, with a significance level of 0.05. Nevertheless, SVM is the best statistically significant method via texture and combined texture-colour features based on Table 4.

Compared to the work by Orchi et al. (2023) in Table 6, RF is competitive against Inception V3 based on F1-score reading even though Orchi et al.'s (2023) work experiments with binary class crop leaf disease. Given a huge dataset, RF has the potential to beat DL classifiers, and the number of instances in each class is more proportionate. In the future, this study recommends improving RF by incorporating an enhanced dynamic weighted function to regularise the voting mechanism to classify.

Either colour features or a combination of texture and colour features are suitable for determining mangosteen maturity stages. This finding is also applicable to previous works, as shown in Table 6 (Al-Mashhadani et al., 2021; Kim et al., 2023; Muñoz et al., 2021; Parashar & Johri, 2024).

Table 6
Summary of previous works performance

Article	Dataset	Type/Classes	Feature	Classifier	F1-score	Recall	Precision	Accuracy (%)
Orchi et al. (2023)	Crop leaf	Leaf disease/2	Colour	Inception V3	0.9652	0.9753	0.9652	98.01
				CNN	0.9132	0.9139	0.9157	93.89
				Resnet50	0.9169	0.9164	0.9241	93.57
				VGG16	0.8977	0.9017	0.9043	87.50
				VGG19	0.8668	0.8694	0.8818	86.70
				RF	0.9771	0.9747	0.9792	97.54
				CART	0.9418	0.9406	0.9424	94.45
				KNN	0.9227	0.9226	0.9196	91.67
				SVM	0.8440	0.8437	0.8438	84.10
				LDA	0.7993	0.7994	0.8001	80.28
				NB	0.5474	0.5995	0.6865	60.09
Al-Mashhadani et al. (2021)	Tomato	Ripeness/4	Colour	AlexNet	1.0000	1.0000	1.0000	99.80
Mohtar et al. (2019)	Mangosteen	Maturity/6	Colour	Inception V3	0.9200	0.9600	0.8800	91.90
Kim et al. (2023)	Tomato	Maturity/4	Colour	DNN	0.9100	0.8900	0.9300	97.00
Miraiei et al. (2021)	Mulberry	Ripeness/2	Colour	DenseNet	N/A	N/A	N/A	98.67
				Inception V3	N/A	N/A	N/A	98.55
				ResNet-18	N/A	N/A	N/A	98.65
				ResNet-50	N/A	N/A	N/A	96.33
Benmoana et al. (2022)	Fuji apple	Ripeness/4	Hyperspectra	CNN	0.9646	0.9651	0.9659	96.50
				ANN	0.8807	0.8837	0.8941	89.50
				SVM	0.9590	0.9607	0.9592	95.93
				kNN	0.9192	0.9186	0.9205	91.18
Gao et al. (2020)	Strawberry	Ripeness/2	Hyperspectra	AlexNet	N/A	N/A	N/A	98.60
Tan et al. (2022)	Beef	Quality/8	Colour	Inception V4	0.6040	0.6130	0.6280	90.00
Muñoz et al. (2021)	Bluberry	Ripeness/5	Colour	CNN	0.9008	0.9092	0.8926	90.74
Azizah et al. (2017)	Mangosteen	Surface defect/2	Colour	CNN	N/A	N/A	N/A	97.50

Table 6 (*continue*)

Article	Dataset	Type/Classes	Feature	Classifier	F1-score	Recall	Precision	Accuracy (%)
Sudana et al. (2020)	Mangosteen	Maturity/7	Colour	CNN	N/A	N/A	N/A	97.10
Phothisonothai and Tantisatirapong (2019)	Mangosteen	Ripeness/3	Texture, Colour	GMM	N/A	N/A	N/A	86.67
Whidhiasih et al. (2012)	Mangosteen	Maturity/3	Texture, Colour	Fuzzy NN	N/A	N/A	N/A	85.00
Riyadi et al. (2020)	Mangosteen	Skin defect/2	Texture, Colour	LDA	N/A	N/A	N/A	94.16
Khojastehnazhand et al. (2010)	Lemon	Quality/3	Colour	Machine Vision	N/A	N/A	N/A	94.04
Liming and Yanchao (2010)	Strawberry	Quality/4	Colour	K-Means cluster	N/A	N/A	N/A	88.80
Damarjati et al. (2017)	Mangosteen	Skin defect/2	Texture	LDA	N/A	N/A	N/A	77.43
Afandi et al. (2021)	Mangosteen	Skin defect/3	Texture	ELM	N/A	N/A	N/A	96.67
Riyadi et al. (2018)	Mangosteen	Skin defect/2	DCT feature	ELM	N/A	N/A	N/A	92.50
Leemans et al. (2002)	Apple	Quality/4	Colour	DA,NN	N/A	N/A	N/A	75.00
Bashir et al. (2023)	Irrigation	Quality/2	Shape	Ensemble ANN	N/A	N/A	N/A	91.44
				ANN (all parameter)	N/A	N/A	N/A	88.36
				ANN(one parameter)	N/A	N/A	N/A	87.84
Joshua et al. (2022)	Environmental	Quality/2	Numerical	BPNN	N/A	N/A	N/A	89.00
				SVM	N/A	N/A	N/A	93.00
				GRNN	N/A	N/A	N/A	97.00
Parashar and Johri (2024)	Apple	Leaf disease/4	Texture, Colour	Inception V3	0.85	0.87	0.84	94.76
Khan, Nauman et al. (2022)	Irrigation saline soil	Quality/2	Shape	Embedded LSTM	N/A	N/A	N/A	78.00
				LSTM	N/A	N/A	N/A	70.00
Khan Faheem et al. (2022)	Soil and fertility	Quality/2	Numerical	LR	N/A	N/A	N/A	92.00
				SVM	N/A	N/A	N/A	94.00
				GNB	N/A	N/A	N/A	96.00
				kNN	N/A	N/A	N/A	93.00

RF also generated the lowest FNR across texture and colour and combined texture-colour feature schemes for almost all six classes, as presented in Figures 10, 12 and 14. The FNR were computed based on FN and TP reading in the confusion matrix for each feature scheme. RF achieved the best statistically significant method among the experimented methods based on the lowest mean rank value analysed using the Friedman Test with a 0.05 significance level. Apparently, RF is the best method to handle FNR-related analysis in mangosteen six multiclasses's maturity stages study. Nevertheless, there is not enough information to claim that other classifiers from agricultural-related studies are able to handle FNR because many studies have not analysed the related metric. In future work to improve weighted function in an enhanced RF algorithm, it is recommended that FNR analysis be incorporated with other machine learners to justify its computational capability. This work is limiting the work among traditional machine learners. Thus, the next work will extend the experimentation using DL instead of comparing it with traditional machine learners.

ACKNOWLEDGMENT

This research was supported by the Ministry of Education (MOE) through the Fundamental Research Grant Scheme FRGS/1/2020/WAB04/UNIMAP/02/3.

REFERENCES

Afandi, M. I., Kurniawan, E., & Wijaya, S. K. (2021). Rapid grading of mangosteen peel defect using Extreme Learning Machine. In *Proceedings of the 2021 International Conference on Computer, Control, Informatics and Its Applications* (pp. 60-65). ACM Publishing. https://doi.org/10.1145/3489088.3489110

Al-Mashhadani, Z., & Chandrasekaran, B. (2021). Ros-based robotic system for tomato disease and ripeness classification using convolutional neural networks. In *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 0420-0427). IEEE Publishing. https://doi.org/10.1109/iemcon53756.2021.9623183

Ashtiani, S. M., Javanmardi, S., Jahanbanifard, M., Martynenko, A. I., & Verbeek, F. (2021). Detection of mulberry ripeness stages using deep learning models. *IEEE Access, 9*, 100380-100394. https://doi.org/10.1109/access.2021.3096550

Azizah, L. M. R., Umayah, S. F., Riyadi, S., Damarjati, C., & Utama, N. A. (2017). Deep learning implementation using convolutional neural network in mangosteen surface defect detection. In *2017 7th IEEE international conference on control system, computing and engineering (ICCSCE)* (pp. 242-246). IEEE Publishing. https://doi.org/10.1109/iccsce41414.2017

Bashir, R. N., Khan, F. A., Khan, A. A., Tausif, M., Abbas, M. Z., Shahid, M. M. A. & Khan N. (2023). Intelligent optimization of Reference Evapotranspiration (ETo) for precision irrigation. *Journal of Computational Science, 69*, Article 102025. https://doi.org/10.1016/j.jocs.2023.102025.

Benmouna, B., García-Mateos, G., Sabzi, S., Fernández-Beltran, R., Parras-Burgos, D., & Molina-Martínez, J. M. (2022). Convolutional neural networks for estimating the ripening state of Fuji Apples using visible and near-infrared spectroscopy. *Food and Bioprocess Technology, 15*, 2226-2236.

Boateng, E. Y., Otoo, J., & Abaye, D. (2020). Basic tenets of classification Algorithms K-Nearest-Neighbor, support vector machine, random forest and neural network: A review. *Journal of Data Analysis and Information Processing*, 8, 341-357. https://doi.org/10.4236/jdaip.2020.84020

Damarjati, C., Riyadi, S., Triyani, W. I., Azizah, L. M., & Hariadi, T. K. (2017). Statistical features extraction of discrete curvelet transform for surface quality evaluation of mangosteen. In *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 236-241). IEEE Publishing. https://doi.org/10.1109/iccsce.2017.8284411

Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., & Han, X. (2020). Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning, *Artificial Intelligence in Agriculture, 4*, 31-38. https://doi.org/10.1016/j.aiia.2020.04.003

Giuntini, R., Holik, F. Park, D. K., Freytes, H., Blank, C., & Sergioli, G. (2023). Quantum-inspired algorithm for direct multiclass classification. *Applied Soft Computing, 134*, Article 109956. https://doi.org/10.1016/j.asoc.2022.109956.

Goel, E., & Abhilasha, E. (2017). Random forest: A review. *International Journal of Advanced Research in Random Forest, 7*(1), 251–257. https://doi.org/10.23956/ijarcsse/v7I1/01113

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 3*, 610–621.

IndexBox. (2024). *Malaysia: Mango and Mangosteen Market 2024*. https://www.indexbox.io/store/malaysia-mangoes-mangosteens-and-guavas-market-report-analysis-and-forecast-to-2020/

Joshua, S. V., Priyadharson, A. S. M., Kannadasan, R., Khan, A. A., Worawat, L., Khan, F. A., Rehman, A. U., & Ali, M. J. (2022). Crop yield prediction using machine learning approaches on a wide spectrum. *Computers, Materials & Continua, 72*(3), 5663-5679. https://doi.org/10.32604/cmc.2022.027178

Khan A. A., Nauman, M. A., Bashir, R. N., Jahangir, R., Alroobaea, R., Binmahfoudh, A., Alsafyani, M., & Wechtaisong, C. (2022). Context aware evapotranspiration (ETs) for saline soils reclamation. *IEEE Access, 10*, 110050-110063. https://doi.org/10.1109/ACCESS.2022.3206009.

Khan, A. A., Faheem, M., Bashir, R. N., Wechtaisong, C., & Abbas, M. Z. (2022). Internet of Things (IoT) assisted context aware fertilizer recommendation. *IEEE Access, 10*, 129505–129519. https://doi.org/10.1109/ACCESS.2022.3228160

Khojastehnazhand, M., Mahmoud, O., & Tabatabaeefar, A. (2010). Development of a lemon sorting system based on colour and size. *African Journal of Plant Science, 4*, 122-127.

Kim, T., Lee, D. H., Kim, K. C., Choi, T., & Yu, J. M. (2023). Tomato maturity estimation using deep neural network. *Applied Sciences, 13(1), Article* 412. https://doi.org/10.3390/app13010412

Leemans, V., Destain, M., & Magein, H. (2002). *On-line fruit grading according to their external quality using machine vision. Biosystems Engineering,* ***88*****(**4), 397-404.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18(8),* Article 2674. https://doi.org/10.3390/s18082674

Liming, X., & Yanchao, Z. (2010). Automated strawberry grading system based on image processing. *Computers and Electronics in Agriculture, 71*(1), 32–39. https://doi.org/10.1016/j.compag.2009.09.013

Mohtar, I. A., Ramli, N. S. S., & Ahmad, Z. (2019). Automatic classification of mangosteen ripening stages using deep learning. In *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (pp. 44-47). IEEE Publishing. https://doi.org/10.1109/aidas47888.2019.8970933

Muñoz, P. B. C., Sorogastua, E. M. F., & Gardini, S. R. P. (2022). Detection and classification of ventura-blueberries in five levels of ripeness from images taken during pre-harvest stage using deep learning techniques. In *2022 IEEE ANDESCON* (pp. 1-6). IEEE Publishing. https://doi.org/10.1109/andescon56260.2022.9989578

Orchi, H., Sadik, M., Khaldoun, M., & Sabir, E. (2023). Automation of crop disease detection through conventional machine learning and deep transfer learning approaches. *Agriculture, 13*, Article 352. https://doi.org/10.3390/agriculture13020352

Pandey, R., Naik, S., & Marfatia, R. (2013). Image processing and machine learning for automated fruit grading system: A technical review. *International Journal of Computer Applications, 81*(16), 29-39. https://doi.org/10.5120/14209-2455

Parashar, N., & Johri, P. (2024). Enhancing apple leaf disease detection: A CNN-based model integrated with image segmentation techniques for precision agriculture. *International Journal of Mathematical, Engineering and Management Sciences, 9*(4), 943-964. https://doi.org/10.33889/IJMEMS.2024.9.4.050

Phothisonothai, M., & Tantisatirapong, S. (2019). Fractal dimension based color texture analysis for mangosteen ripeness grading. In *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (pp. 1-2). IEEE Publishing. https://doi.org/10.1109/ispacs48206.2019.8986398

Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63. https://doi.org/10.48550/arXiv.2010.16061

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports, 14*, Article 6086. https://doi.org/10.1038/s41598-024-56706-x

Riyadi, S., Azizah, L. M. R., Damarjati, C., & Hariadi, T. K. (2018). Evaluation of mangosteen surface quality using discrete curvelet transform. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 475-479). IEEE Publishing. https:// 10.1109/ictc.2018.8539577

Riyadi, S., Ratiwi, A. M. A., Damarjati, C., Hariadi, T. K., Prabasari, I., & Utama, N. A. (2020). Classification of mangosteen surface quality using principal component analysis. *Emerging Information Science and Technology, 1*(1), 33-43. https://doi.org/10.18196/eist.115

Rushing, B. (2022). No free theory choice from machine learning. *Synthese, 200*, Article 414. https://doi.org/10.1007/s11229-022-03901-w

Sudana, O., Bayupati, I. P. A. & Yudiana, D. G. (2020). Classification of maturity level of the mangosteen using the convolutional neural network (CNN) method. *International Journal of Advanced Science and Technology (IJAST), 135*(2020), 37-48. https://doi.org/10.33832/ijast.2020.135.04

Tan, W. K., Husin, Z., & Ismail, M. A. H. (2020). Feasibility study of beef quality assessment using computer vision and Deep Neural Network (DNN) algorithm. In *2020 8th International Conference on Information Technology and Multimedia (ICIMU)* (pp. 243-246). IEEE Publishing. https://doi.org/10.1109/icimu49871.2020

Thammastitkul, A., & Klayjumlang, T. (2021). Mangosteen quality grading for export markets using digital image processing techniques. *International Journal on Advanced Science Engineering Information Technology, 11*(6), 2452–2458. https://doi.org/10.18517/ijaseit.11.6.14003

Whidhiasih, R. N., Guritman, S., & Supriyo, P. T. (2012). Identification of mangosteen stage maturity on colour based using fuzzy neural network. *Jurnal Teknologi Industri Pertanian, 22*(2), 82-91.