

Towards Smart Agriculture Through Auto-Quality Checker And Grader Of Mango Fruits Using Ensemble Machine Learning Model

*Kalyani C. Waghmare¹, Anupama G. Phakatkar², Rekha A. Kulkarni³, Mukta S. Takalikar⁴

¹Assistant Professor, Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, Maharashtra, India, kcwaghmare@pict.edu ¹

²Assistant Professor, Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, Maharashtra, India, agphakarkar@pict.edu ²

³Assistant Professor, Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, Maharashtra, India, rakulkarni@pict.edu ³

⁴Associate Professor, Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune, Maharashtra, India, mstakalikar@pict.edu ⁴

ABSTRACT:

A variety of Picture processing and ML technologies currently used in the systems that classify mango quality. Traditional algorithms usually work individually and aim to identify patterns of functions in data records. However, all technologies have their own strengths and limitations. This study introduces an ensemble learning frame that integrates several ML models to improve prediction accuracy. The process begins with extracting external features of mangoes with different image processing techniques. These functions combine with sensor-derived weight values to form a data record. Several ML algorithms are tested on this data record to identify the most effective ones. Next, ensemble learning strategies such as baking, boosting, and stacking are used. Experimental results show that image processing methods significantly reduce errors. Furthermore, model assessments show that the stacking ensemble combined various basic and meta-learners with accuracy, recall, F1 scores, and accuracy values with 0.9855, 0.9901, 0.9876 or 0.9863. These results confirm the effectiveness and reliability of the proposed approach for mango quality classification.

Keywords: Mango grading, feature extraction, ML, Vietnam agriculture, fruit classification.

Introduction

Rapid advances technology the agricultural sector paved the way for the integration of AI and ML in traditional agricultural practices, leading to the concept of intellectual agriculture. One of the most important applications of intelligent agriculture is the automated assessment of fruit quality, which plays a key role in improving follow-up processes, reducing human error, and fulfilling market standards. Under tropical fruits, mangoes have great economic and commercial value in the global market. However, traditional methods of high quality testing and stamina are often subjective, time-consuming and

inconsistent. This has led to an increasing demand for mangoes based on a variety of quality attributes, such as intelligent and automated systems, size, shape, color, and texture.

This study examines the development of automated mango quality testers and graders with an ensemble approach for machine learning. The purpose is to provide a more reliable and more accurate classification mechanism by using the combined strength of several ML algorithms, such as sack, increase, and stacking of this system. The framework uses extended image processing techniques to extract visual features from mango images and is integrated into sensor data to form a comprehensive data set for training and evaluation. The proposed ensemble-based system not only improves classification performance, but also contributes to the scalability and robustness of actual agricultural applications. The purpose of this study is to promote manual quality inspection and fully automated and intelligent decision-making - excellent decisions in fruit quality control and ultimately efficiency and sustainability in modern agriculture.

Related Work

At peak times, mangoes are harvested on farms and distributed to various markets. However, due to regional differences in ground composition and climate, quality will vary from region to region. Despite this natural variation, many markets need uniform, high quality fruits, especially internationally, and mango classification must take important steps to ensure consistency. Traditionally, mango classification is performed manually and is strongly based on the experience of farmers. This method is common, but often leads to consistency and unreliable classification. However, manual approach increases both the cost of equipment and the work requirements. Furthermore, the costs of agricultural production have increased significantly. To address these challenges, commercially available machines have been developed to automate the agricultural classification process.

Previous studies have investigated various methods of fruit evaluation. For example, some studies used electronic nasal instruments to assess fruit maturation [5], while another [6] proposed. By analyzing Spectral data to estimate the intensity and using a fuzzy classifier, this method allows accurate classification of fruit maturity despite deviations. However, these non-bio-based approaches are subject to the limitations of large-scale applications due to complexity and high work requirements. To overcome these drawbacks, a new study examined a machine vision-based fruit classification system that provides greater accuracy and efficiency. Machine learning algorithms known for their adaptability are often used in computer vision. The system categorized mangoes in four classes, taking into account the removal and value of the market, reaching 87% accuracy and 100% repeatability. Further studies have shown the effectiveness of ML in mango classification. In [8], the researchers classified mangoes using visible images and form analysis, achieving 98.3% accuracy in discriminatory analysis and 100% in support vector machine (SVM). Using a cylindrical approach, a strong correlation of 94% was found between the estimated weight and the actual weight.

The results confirmed the high accuracy of the system in grain classification. To address this, researchers turned deep learning (DL) models (DL). For example, Cao et al.[11] Introduced Light Net, a light CNN architecture for fresh Zizania evaluation. With the parallel processing design, Light Net achieved 95.62%

accuracy per image in the Zizania classification and 99.31% accuracy in the Apple classification, reducing arithmetic requirements. Another study [12] used CNNs with transfer learning for seed quality classification. With Google Net, the system achieved 95% accuracy, significantly outperforming traditional SURF+SVM methods, achieving only 79.2%. Despite its accuracy, the DL model is mathematically intensive and inappropriate for real-time classification on limited devices. In addition, ML algorithms are less comprehensive in the study [8-11], but have been implemented individually, limiting the ability to use the combined strengths of several models for optimal performance. Ultimately, the most effective and reliable models were integrated throughout the system.

The most important contributions to this research are explained below.

1. The proposed stack ensemble approach uses the compound strength of several algorithms of several machine learning, significantly improving the prediction accuracy of the system.
2. This method was successfully implemented for mango classification, improving the efficiency of mango supply and distribution chains.
3. This approach is adaptive and can be easily extended to classify other agricultural products such as sweet potatoes, tomatoes and other products, and to support the development of intelligent technologies and tools in modern agriculture

Proposed Method

3.1 Structure of the Mango Sorting System

From the Figure 1, mango evaluation system consists of several integrated components that process raw mangoes and process output. First of all, farmers will manually inspect the mangoes to remove anyone who is too small or seriously damaged. The selected mangoes are placed on a conveyor belt (Fig. 1a) and transported to an image processing chamber (Fig. 1a). Mango then passes through a load system equipped with a load cell. Sensor data is filtered using a Kalman filter and then sent to the central processing unit for further evaluation. Although this study does not focus on detailed processing of load cell signals, the resulting weight data is combined with external functions to form the input of the classification model.



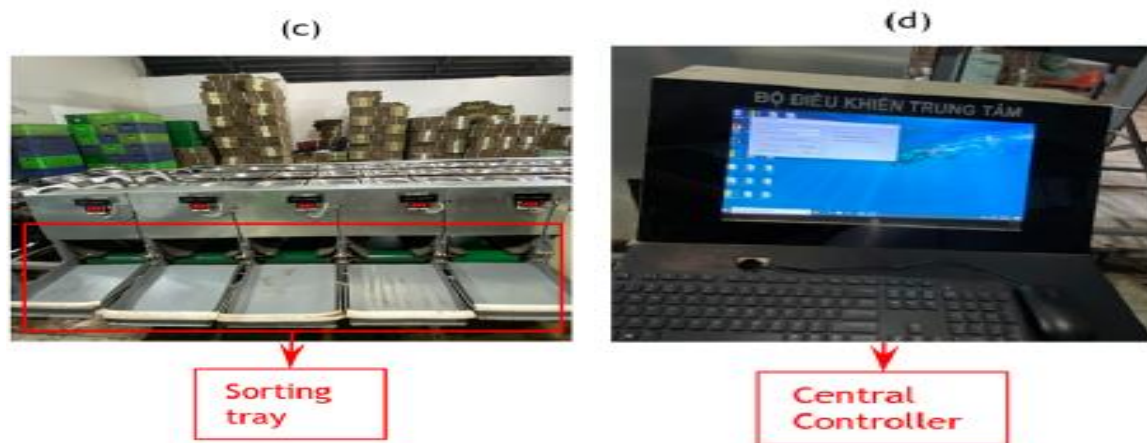


Figure 1 Mango-grading system

Following distinctive extraction and data collection moves the mango into the sorting shell area via another sponsor (Figure 1b). At the sort station (Figure 1c), the classification system leads all mangoes based on the category on the corresponding tray. The entire sorting process is controlled and adjusted by a central controller (Figure 1d) to complete the automated evaluation workflow.

The classification process begins with extracting external functions by an algorithm run on a central processor. As soon as the data is processed, it is fed into a classification model that predicts all mango quality categories. Based on the predicted categories, the corresponding sorting mechanism is activated, ensuring that each mango is accurately classified as shown in Figure 2.

3.2. Extracting External Features of Mangoes Using Image Processing

This section describes the Picture processing algorithms in which external properties are extracted. This process involves several important steps. First, mangoes are isolated from the background by image segmentation.

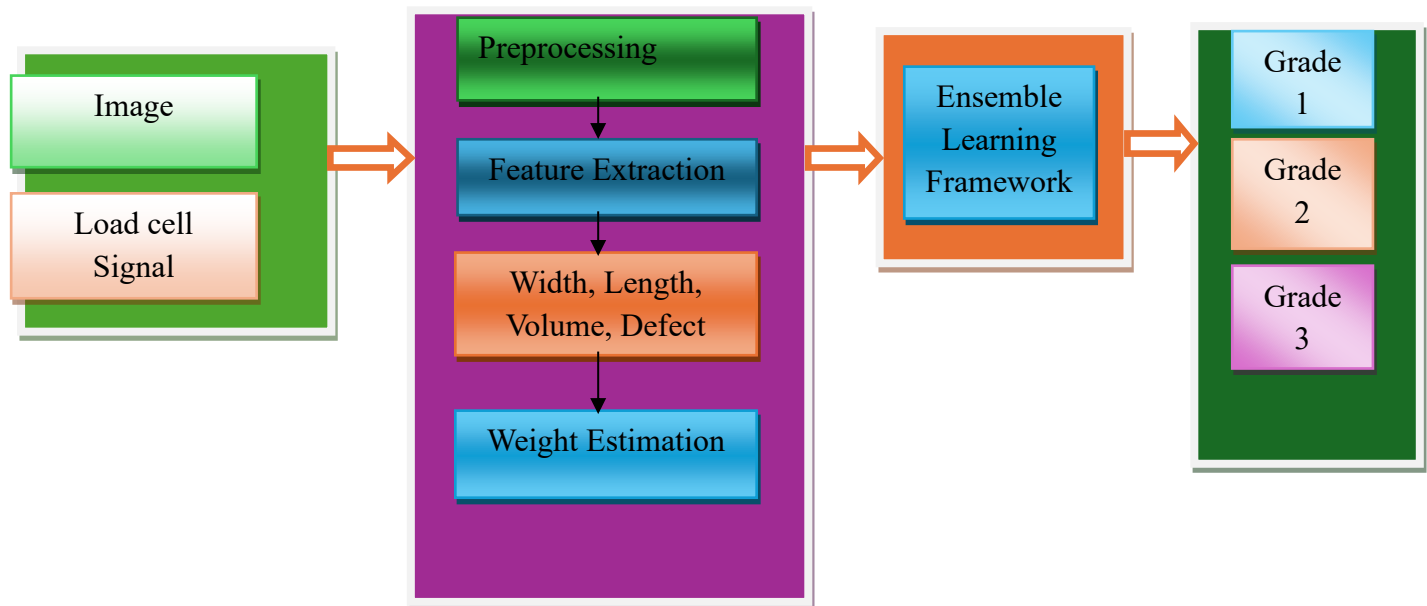


Figure 2 Mango-quality grading process

3.2.1. Mango Metamerism

Explained in the earlier section, the mango where brought into commercial commercial along with the scooter conveyor. In this setup, the mangoes are initially isolated from the background. Recent research has investigated various object segmentation algorithms, and deep learning (DL) methods have proven to be the most effective. However, DL models typically require important arithmetic resources, which slows down the processing of classification systems. To overcome this limitation, there have been traditional image processing methods with fewer computing resources used to segment mangoes. This method is to use a fixed threshold for grayscale images to distinguish objects from background, as explained in [14]. Despite its simplicity, this approach is inflexible when mango color variations occur. Therefore, in this case, using HSV color space for segment-mangos [4] consists of more effective segmentation technology.

This setup plays an important role in the production of high quality input images with minimal noise and provides ideal conditions for effective image binarization. As shown in Figure 3, the green color of the mango is in stark contrast to the background, enhancing segmentation.

Original Image HSV Color Image

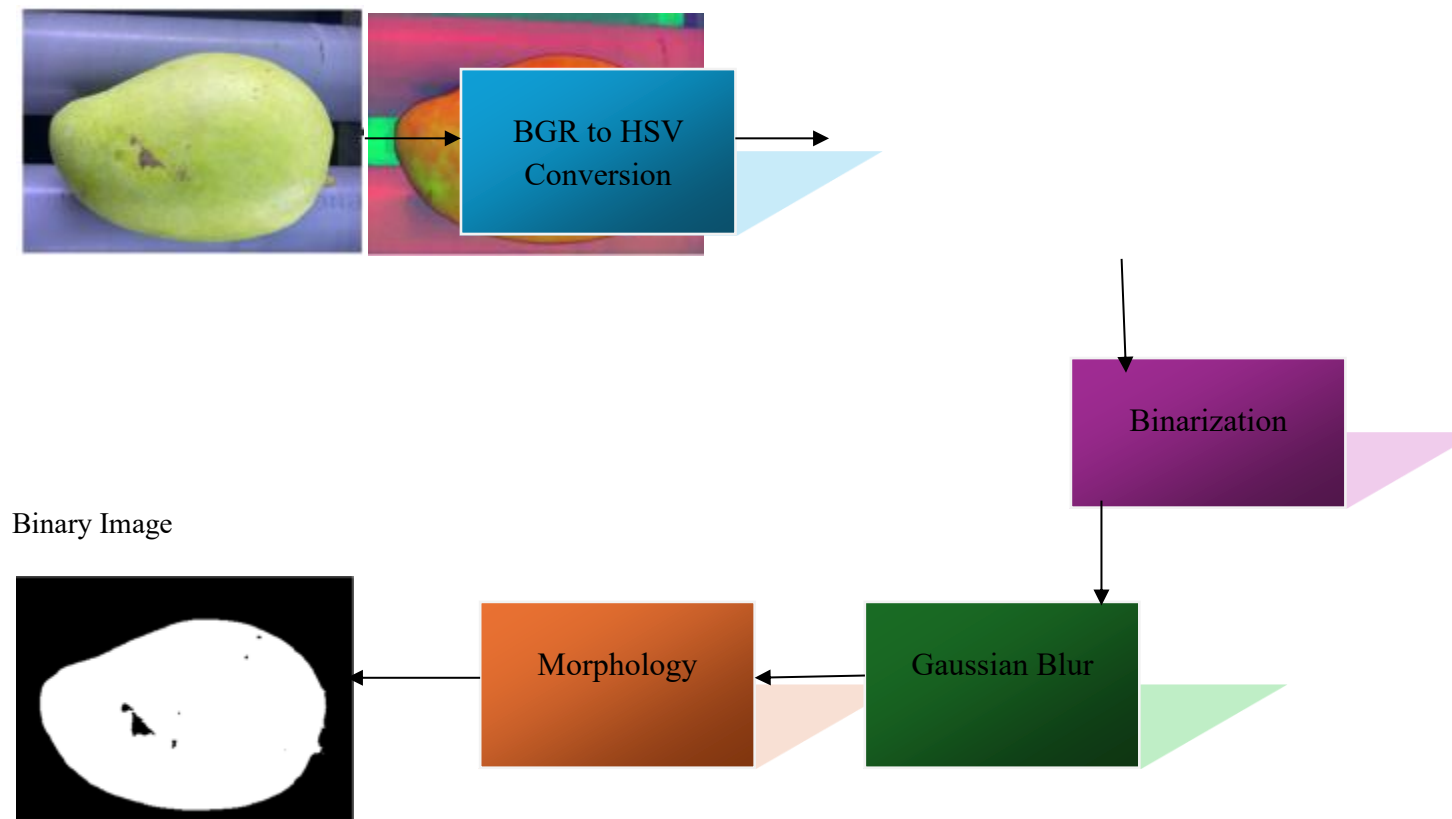


Figure 3 Mango-segmentation method

During specific region of the HSV spectrum defined, with $T_{\min}(h, s, v)$ set from (30, 15, 17) and $t_{\max}(h, s, v)$ to (80, 255, 255). Pixels within this region are assigned a value of 255 to the binary image. This indicates the mango region, but pixels representing the background or defective region are given zeros. To improve the binary image, the image is smoothed using Gaussian blur filter with 3-3 kernels, followed by an expansion process to reduce noise.

3.2.2. Extraction of External Mango Features

As soon as the mangoes are segmented using a bisecting process from the background, the resulting image is processed with different image processing techniques to extract external features. The selection of input functions is based on the results of a study in [15] that focused on fruit classification. Properties used in this study include this analysis, along with biological properties unique to mango and surface defects. These characteristics serve, as shown in Fig 4.

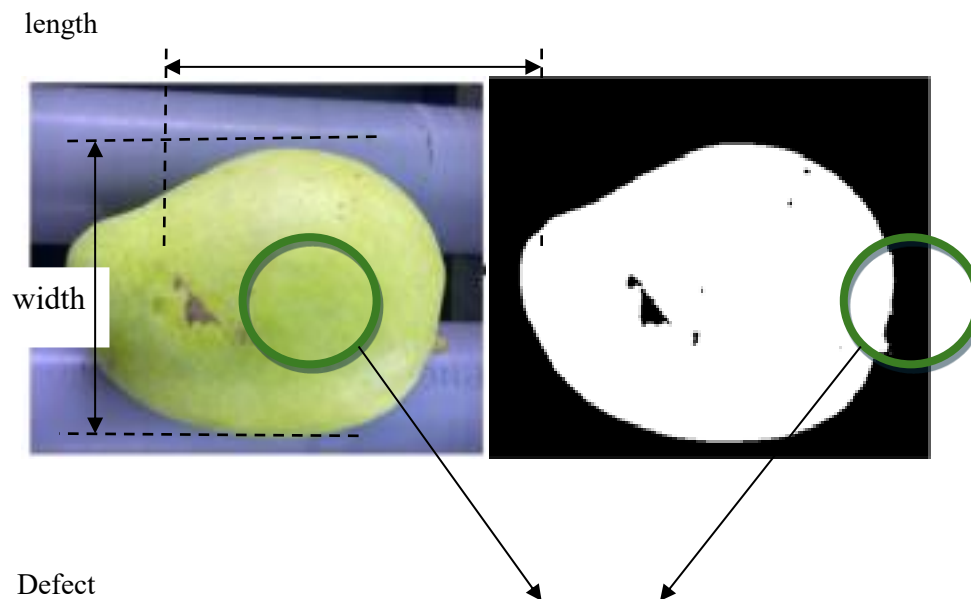


Figure 4 Extraction of External Mango Features

As mangoes move along the role of the conveyor, their functions are recorded using a rectangular bounding box. However, due to the minimal variation, the mango length remains consistent across several images. Due to the irregular shape of the fruit, the maximum width (width max) is passed through knowledge of classification, and the total area of this defect is passed through the surface of the fruit.

While the mango is turning recognize the total area of the defect. After extracting external functions, these are integrated into the weight measurements collected by sensor.

$$W_{\text{actual}} = kW_{\text{pixel}}(1)$$

In this context, w_{pixels} represent the width of a mango in pixels. where W_{actual} refers to the actual width transformation coefficient. Maintain measurement, the coefficient K must be approximated. As the mango approaches the camera, pixels are displayed and the number of pixels increases. As a result, increasing this number of pixels to K also increases the actual measurement. Therefore, to take this variation into account, we need to define K as a function.

$$W_{\text{avg}} = 1/n \sum_{i=1}^n W_i(2)$$

3.3 Information Collection

This study [17], large and diverse data records significantly improve the output of the model. To support this, structured data records were collected at several harvests. Data record contains six columns. A total of 1,300 mango samples (see Figure 5) were used to create a training data set. Each characteristic was measured several times (n time) and the mean value was calculated using Equation . All mango rehearsals

were featured by agricultural experts according to the Vietnam Gap (Vietnamese excellent agricultural practices) standards to ensure accurate classification. Each measured feature was documented and the corresponding error was calculated to determine the deviations between the actual system. These error analysis will be further discussed in the Experimental section.



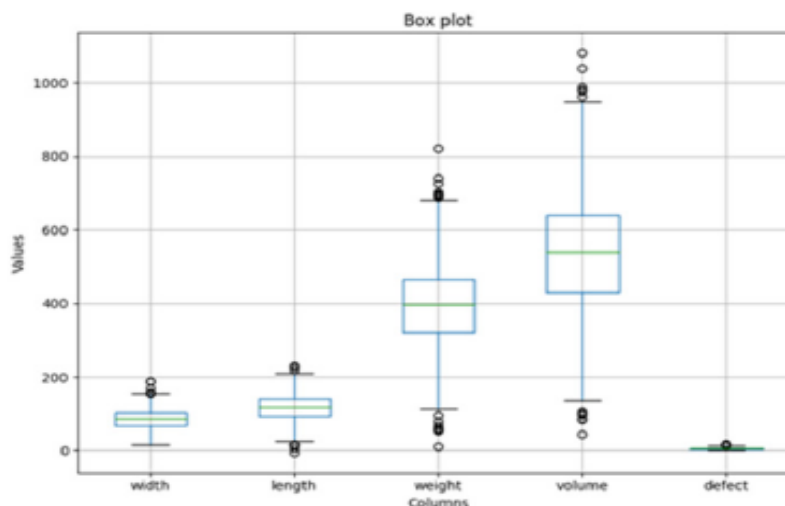
Figure 5 Mango samples

Level of study	Width (cm)	Length (cm)	Volume (mL)	Weight (g)	Defect (cm ²)
1	9-11	14.1-16	651-800	451-700	0-3
2	8-9	12.1-14	401-650	250-450	3-5
3	6-8	10-12	250-400	100-250	>5

Table 1Mango-quality standards

3.3.2 Data Preprocessing

To improve this, standardization of the data is used as an important pre-processing step to minimize the effects of different characteristic scales. This technology is very importantNormalization adapts the domains and scales of characteristics and prevents features with large numerical values that disproportionately affect the learning process. This leads to more balanced and consistent model training. Furthermore, normalization increases the speed of convergence of the optimization algorithm, allowing the optimal solution to be realized more efficiently. By promoting data uniformity and stability, normalization enhances model performance and accuracy, leading to a more reliable classification of mango quality. Figure 6 shows the relationships between the characteristics of data records.



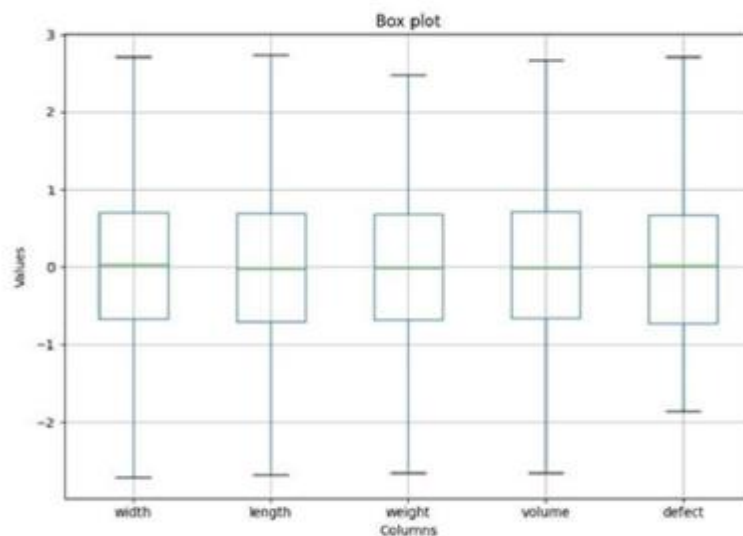


Figure 6 Distribution of features in the information set: (a) original information, (b) processed information

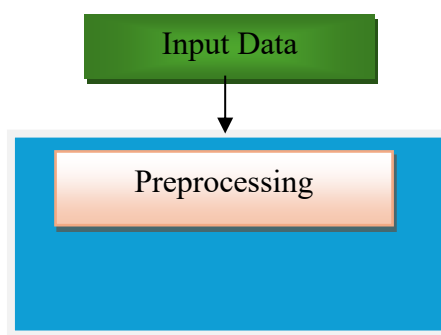
Using standard scaling methods, we normalized the data records to ensure that all characteristics were scaled consistently and that all distortions caused by the original size differences were removed. This standardization technique has a mean zero and standard deviation to adapt each characteristic, effectively eliminating differences due to different characteristics. The equation for the standard scale is described in Equation (3).

$$z = (x - \mu) / \sigma \quad (3)$$

In this formula, x represents the original feature value, μ is the mean of all values for that feature, σ denotes the standard deviation, and z is the resulting standardized value.

3.3.3. Suggested Ensemble Learning Approach

This procedure begins with preprocessing the data and converts the label values in the string sequence into numeric format. Next, all features recognize and remove outliers to minimize data noise. Each model is evaluated independently using K-Times cross-validation techniques to achieve more reliable and more accurate results. Based on this evaluation, a top performance algorithm is selected for further improvements through hyperparameter tuning to improve performance.



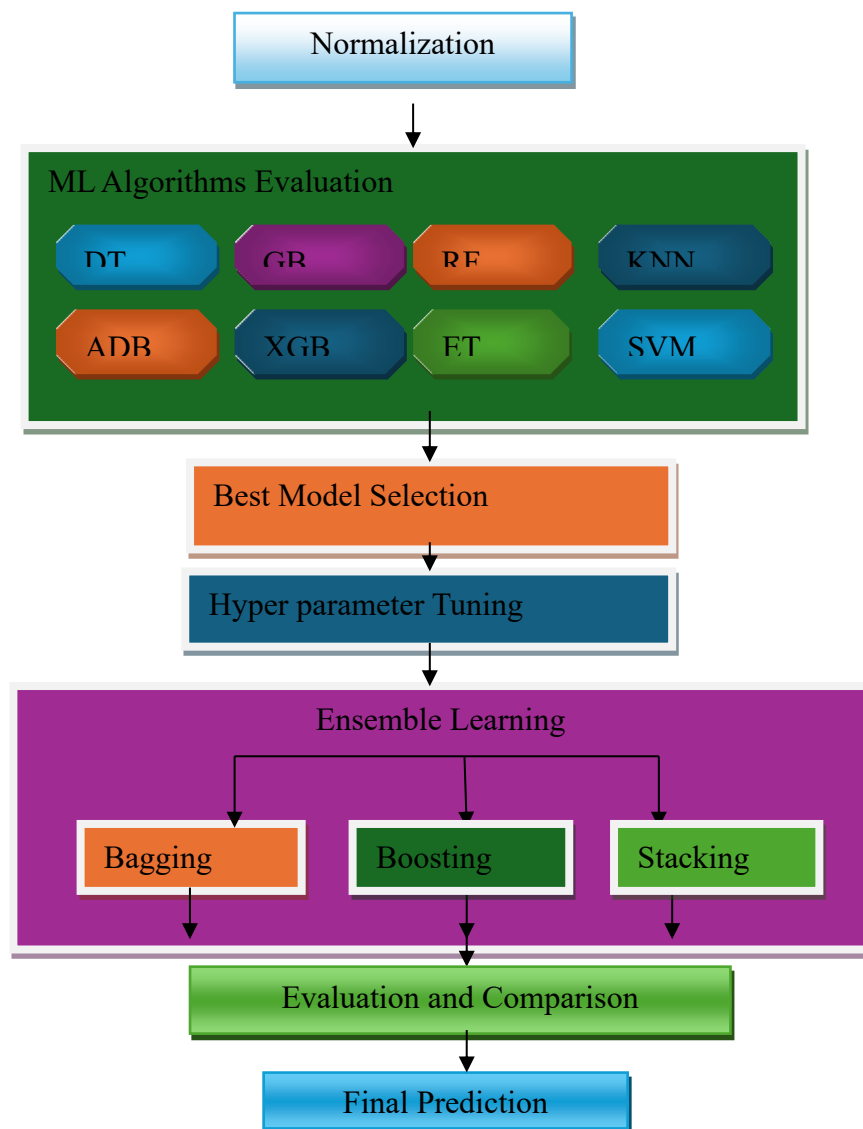


Figure 7Design of the suggested ensemble-learning technique

In this study, K-Fold cross-validation has been selected for the efficiency and widespread use demonstrated in similar studies [24]. This technology divides data records into K-equal sub-amounts or "folded". The model is trained, one times more for training, and the remaining folds are used for verification each time. This allows each subgroup to be used for both training and validation, allowing for a thorough and round performance assessment. The final evaluation metric is calculated by measuring the results of all k characters and creating a reliable model output. Hyperparameter adjustments are also required to optimize machine learning models. Common tuning methods used in the current study include every suitable for different data types and scenarios [22, 23]. The overall process shown in Fig 7

Result

An image processing system was used to extract external features of the mango. Important attributes such as length, width, volume, and surface defects play an important role. By image processing techniques, these properties can be quantitatively analyzed for effective mango classification. To assess the accuracy of the characteristic estimation, the predicted values obtained from the system are compared to actual measurements. The performance of the image processing algorithm is evaluated using two evaluation metrics. These metrics are defined in equations (7) and (8).

$$\text{MAE} = 1/n \sum_{i=1}^n |y - \hat{y}| \quad (4)$$

$$\text{RMSR} = \sqrt{1/n \sum_{i=1}^n (y - \hat{y})^2} \quad (5)$$

A hyperparameter voting is then performed that improves accuracy and reliability. After this step, the three algorithms with the best performance behavior of the ensemble learning method include the order of individual models for recording the roof, including drag, increase of individual models, increase of individual models, etc. The room rooms are exposed to the back model rooms for roof movement for breakfast room work. Stacking is the highest accuracy of them, indicating superior robustness compared to other models. According to Table 2, XGboost surpasses other individual algorithms. As a result, it is chosen as the base model for absorption and increase techniques. In looting, some models are trained independently of each other. With different sub-quantities of training data, the majority are combined. This method of iterative correction often leads to improved prediction accuracy.

Technique	Target	Model Scale	Accuracy	FPS
LDA [28]	Pomegranate	-	91%	60
XGBoost	Mango	2 M	98%	34
Random Forest	Mango	2.2 M	97%	32
Extra Tree	Mango	4.9 M	97%	32
Proposed Method	Mango	5.2 M	99%	28

CNN(eight layers) [30]	Apple	-	93%	19
---------------------------	-------	---	-----	----

Table 2Summary of research results on performance and speed in fruit classification tasks

System measured in frames per second (FPS) on average over 2000 attempts to obtain the final value. This table is organized in descending order of FPS and includes traditional image processing techniques used for object classification, ML, and DL. According to Table 7, LDA [30] is the fastest way to reach 58 fps, probably because it does not include the learning phase. ML and DL methods show slower processing speeds due to arithmetic requirements of the learning process. For example, the CNN model of the [31] process is a 18 fps. In general, ML methods exceed DL in terms of speed, as they become easier with parameters. The EL method is due to the larger model size of 5.2 million parameters compared to XGB, RF, and ET with 2 million, 2.2 million, and 4.9 million respectively. In terms of classification accuracy, ML and DL methods exceed LDA, as intelligent algorithms can extract properties and adapt to a variety of conditions.

Conclusion

This study provides an ensemble learning approach to classifying mango quality. Experimental results showed that the stacking model combined model that provided the good performance. By integrating the strengths of different algorithms for machine learning, the model effectively established relationships within the input data, leading to reliable final predictions. The proposed method greatly improves the accuracy and efficiency of mango quality classification. Practicality provides valuable solutions for agricultural researchers, particularly when managing challenges related to extracting image features and establishing reliable predictive models. Furthermore, this approach is not specifically for mango classification. It lays the foundation for the development of other fruit classification systems, such as sweet potatoes and grapefruit. Apart from agriculture, this ensemble learning frame can be extended to other fields, including mechanical engineering and economics, to solve complex forecasting tasks. A common method is to increase the number of models for each layer to adjust to your specific application. This provides a versatile reference for researchers in various fields.

Reference

1. Momin, M.; Rahman, M.; Sultana, M.; Igathinathane, C.; Ziauddin, A.; Grift, T. Geometry-based mass grading of mango fruits using image processing. *Inf. Process. Agric.* 2017, 4, 150–160.
2. Gururaj, N.; Vinod, V.; Vijayakumar, K. Deep grading of mangoes using convolutional neural network and computer vision. *Multimed. Tools Appl.* 2023, 82, 39525–39550.

3. Thong, N.D.; Thinh, N.T.; Cong, H.T. Mango sorting mechanical system combines image processing. In Proceedings of the 2019 7th International Conference on Control, Mechatronics and Automation (ICCMA), Delft, Netherlands, 6–8 November 2019; pp. 333–341.
4. Trieu, N.M.; Thinh, N.T. A study of combining knn and ann for classifying dragon fruits automatically. *J. Image Graph.* 2022, 10, 28–35.
5. Brezmes, J.; Fructuoso, M.; Llobet, E.; Vilanova, X.; Recasens, I.; Orts, J.; Saiz, G.; Correig, X. Evaluation of an electronic nose to assess fruit ripeness. *IEEE Sens. J.* 2005, 5, 97–108.
6. Matteoli, S.; Diani, M.; Massai, R.; Corsini, G.; Remorini, D. A spectroscopy-based approach for automated nondestructive maturity grading of peach fruits. *IEEE Sens. J.* 2015, 15, 5455–5464.
7. Nandi, C.S.; Tudu, B.; Koley, C. A machine vision technique for grading of harvested mangoes based on maturity and quality. *IEEE Sens. J.* 2016, 16, 6387–6396.
8. Sa'ad, F.; Ibrahim, M.; Shakaff, A.; Zakaria, A.; Abdullah, M. Shape and weight grading of mangoes using visible imaging. *Comput. Electron. Agric.* 2015, 115, 51–56. [CrossRef]
9. Schulze, K.; Nagle, M.; Spreer, W.; Mahayothee, B.; Müller, J. Development and assessment of different modeling approaches for size-mass estimation of mango fruits (*Mangifera indica* L., cv. 'NamDokmai'). *Comput. Electron. Agric.* 2015, 114, 269–276.
10. Mittal, S.; Dutta, M.K.; Issac, A. Non-destructive image processing based system for assessment of rice quality and defects for classification according to inferred commercial value. *Measurement* 2019, 148, 106969.
11. Cao, J.; Sun, T.; Zhang, W.; Zhong, M.; Huang, B.; Zhou, G.; Chai, X. An automated zizania quality grading method based on deep classification model. *Comput. Electron. Agric.* 2021, 183, 106004.
12. Huang, S.; Fan, X.; Sun, L.; Shen, Y.; Suo, X. Research on classification method of maize seed defect based on machine vision. *J. Sens.* 2019, 2019, 2716975. [CrossRef]
13. Pérez-Borrero, I.; Marín-Santos, D.; Gegúndez-Arias, M.E.; Cortés-Ancos, E. A fast and accurate deep learning method for strawberry instance segmentation. *Comput. Electron. Agric.* 2020, 178, 105736. [CrossRef]
14. Li, Z.; Yin, C.; Zhang, X. Crack Segmentation Extraction and Parameter Calculation of Asphalt Pavement Based on Image Processing. *Sensors* 2023, 23, 9161.
15. Ghazal, S.; Qureshi, W.S.; Khan, U.S.; Iqbal, J.; Rashid, N.; Tiwana, M.I. Analysis of visual features and classifiers for Fruit classification problem. *Comput. Electron. Agric.* 2021, 187, 106267.

16. Chithra, P.; Henila, M. Apple fruit sorting using novel thresholding and area calculation algorithms. *Soft Comput.* 2021, 25, 431–445. [CrossRef]
17. Behera, S.K.; Rath, A.K.; Sethy, P.K. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. *Inf. Process. Agric.* 2021, 8, 244–250.
18. T.K., B.; Annavarapu, C.S.R.; Bablani, A. Machine learning algorithms for social media analysis: A survey. *Comput. Sci. Rev.* 2021, 40, 100395. [CrossRef]
19. Sen, P.C.; Hajra, M.; Ghosh, M. Supervised classification algorithms in machine learning: A survey and review. In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*; Springer: Singapore, 2020; pp. 99–111.
20. Zhang, S. Cost-sensitive KNN classification. *Neurocomputing* 2020, 391, 234–242.
21. Ghiasi, M.M.; Zendehboudi, S. Application of decision tree-based ensemble learning in the classification of breast cancer. *Comput. Biol. Med.* 2021, 128, 104089. [CrossRef]
22. Ileberi, E.; Sun, Y.; Wang, Z. Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access* 2021, 9, 165286–165294.
23. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* 2021, 54, 1937–1967. [CrossRef]
24. Qiu, Y.; Zhou, J.; Khandelwal, M.; Yang, H.; Yang, P.; Li, C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Eng. Comput.* 2021, 38, 4145–4162.
25. Cui, S.; Yin, Y.; Wang, D.; Li, Z.; Wang, Y. A stacking-based ensemble learning method for earthquake casualty prediction. *Appl. Soft Comput.* 2021, 101, 107038.
26. Wei, H.; Chen, W.; Zhu, L.; Chu, X.; Liu, H.; Mu, Y.; Ma, Z. Improved Lightweight Mango Sorting Model Based on Visualization. *Agriculture* 2022, 12, 1467.
27. Long, N.T.M.; Thinh, N.T. Using machine learning to grade the mango's quality based on external features captured by vision system. *Appl. Sci.* 2020, 10, 5775.
28. Wang, S.-H.; Chen, Y. Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique. *Multimed. Tools Appl.* 2020, 79, 15117–15133.
29. Trieu, N.M.; Thinh, N.T. Quality classification of dragon fruits based on external performance using a convolutional neural network. *Appl. Sci.* 2021, 11, 10558.

30. Blasco, J.; Cubero, S.; Gómez-Sanchís, J.; Mira, P.; Moltó, E. Development of a machine for the automatic sorting of pomegranate(*Punica granatum*) arils based on computer vision. *J. Food Eng.* 2009, 90, 27–34.
31. Fan, S.; Li, J.; Zhang, Y.; Tian, X.; Wang, Q.; He, X.; Zhang, C.; Huang, W. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* 2020, 286, 110102.