International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A MACHINE LEARNING APPROACH TO MULTI-CLASS FRUIT RIPENESS DETECTION USING STRUCTURED PHYSICOCHEMICAL DATA

Chandan Prasad[1], Dr. Shailesh Kumar[2], Dr. Sunil Damodar Rathod[3]

[1]Computer Engineering Department, JJT University, Jhunjhunu, Rajasthan, India

[2]Computer Science and Engineering, Gopalan College of Engineering and Management, Bengaluru, Karnataka, India

[3]Indira College of Engineering and Management, Savitribai Phule Pune University, Pune, India

Abstract

Fruit ripeness prediction has an essential role in post-harvest quality management and food supply chain optimization. The majority of current methods rely on image-based or sensordriven systems, which are expensive, complicated, and not normally applicable to multi-class classification. Here, this research presents a low-cost and explainable machine learning architecture for orange ripeness prediction with structured physicochemical and categorical features. The dataset consisted of 241 samples with ripeness labeled into three classes: Unripe, Semi-Ripe, and Ripe. Preprocessing consisted of one-hot encoding, normalization, and SMOTE for class imbalance. Five machine learning algorithms—Logistic Regression, Random Forest, XGBoost, SVM, and MLP—were trained and compared. The best accuracy (75%) was provided by the Random Forest model, followed by XGBoost with an accuracy of 74%, having better class-wise prediction. Feature importance analysis indicated that the three most influential predictors of ripeness were Softness, Brix (Sweetness), and Harvest Time. The results also point to the usefulness of engineered tabular data in ripeness classification and to the fact that deep learning algorithms may be equalled by hand-designed machine learning systems when suitably trained with quality features. This approach is scalable and has the potential for deployment in actual agriculture environments for fruit classification.

Keywords: Ripeness detection, physiochemical data, Machine learning models, multiclass classification, post-harvest automation.

1. Introduction

Fruit maturity is the single most important factor to determine market worth, consumer acceptability, dietary quality, and shelf life. Proper identification of the optimal maturity level ensures that fruits are harvested properly, stored in due time, and supplied to markets before spoilage. Historically, ripeness grading has been based on manual inspection techniques that evaluate visual and tactile characteristics like color, softness, and surface blemishes.

Although this method is commonly used, it is subjective, time-consuming, and prone to inconsistency, particularly for large-scale agricultural production. Due to these shortcomings, researchers have sought different technological solutions for the automated detection of ripeness. The majority of these are based on image-based methods, such as RGB imaging, hyperspectral imaging, and computer vision based on deep learning, to measure and analyze external features as shown in Fig.1.

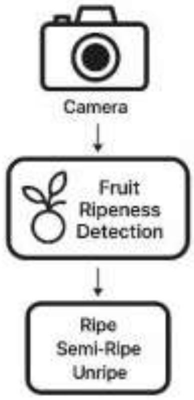


Fig.1: Fruit Ripeness detection

Though effective, such systems are generally demanding in terms of expensive imaging setups, specific lighting conditions, and large computational requirements, hence less viable for field-level or small-scale applications. Furthermore, most such studies address binary classification (ripe or unripe) while neglecting the middle stage, such as semi-ripe, that is significant in making finer decision points during harvesting, processing, and storage. Structured physicochemical traits, like Brix (sugar), pH, softness, and date of picking, are objective, quantifiable, and conveniently found in the routines of fruit manufacturing and quality inspection. Such traits are nevertheless poorly utilized within ripeness recognition models, with fewer attempts having been made at assessing their overall use in multi-class classification setups. This study fills these shortcomings through the creation of a machine learning pipeline that uses non-image, structured data for orange ripeness classification into three categories: Unripe, Semi-Ripe, and Ripe. The method presents a cost-effective, interpretable, and scalable solution appropriate for industrial as well as field-level applications.

1.1 Novelty and contribution of this studyct

The innovation of this research is its application of organized, physicochemical, and categorical fruit attributes—e.g., softness, Brix (sweetness), pH, weight, time of harvest, color, and blemishes—to the multi-class classification of orange maturity. In contrast to most current methods, which heavily depend on image-based or hyperspectral data, this work shows that tabular, sensor-based data is also an effective input for machine learning models, providing a

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

cheaper and more accessible alternative for the prediction of ripeness. In addition, the task of detecting ripeness is tackled in a three-class setup—Unripe, Semi-Ripe, and Ripe—providing finer granularity in the classification than binary methods typically seen in the literature.

This study offers a complete pipeline of data preprocessing, class mapping, encoding, normalization, and SMOTE-based class balancing in addressing the inherent class skew in ripeness distribution. Comparative assessment of five machine learning algorithms. Along with performance assessment, the research also offers interpretability in terms of confusion matrix analysis and feature importance ranking, thus marking softness, Brix, and harvest time as primary determinants of ripeness. In general, the research presents a realizable and interpretable machine learning solution that can be applied in real-world post-harvest citrus fruit classification systems. Its contributions solve significant knowledge gaps in current literature, mainly the under exploitation of tabular physicochemical information and the scarcity of multi-class models in ripeness detection. The suggested methodology is scalable and can be applied to other similar agricultural datasets, thus being a useful tool for supply chain decision-making, quality control, and smart farming applications.

2. Literature Review

2.1 Ripeness Detection Using Color and texture analysis

Fruit ripeness detection is a critical field of research in agricultural automation because of its direct link to post-harvest quality control, marketing, and food safety. In the past, ripeness is determined manually on the basis of visual characteristics such as color, size, and texture. It is not only time-consuming but also inconsistent as it is subjective to human eyes. As a response, various machine learning and computer vision-based solutions have been developed to classify fruit ripeness automatically using physical appearance and internal biochemical change. [1] point to the increasing dependence on machine learning and deep learning algorithms for automating ripeness classification. Their work highlights that deep learning models, specifically, obviate the necessity of manually engineered features, providing flexibility in different types of fruits. They highlight the prevalence of color and texture features in classification problems and discuss several automated approaches that seek to enhance reliability and scalability. Along the same line, [2] introduced a machine vision system for the classification of oil palm fruit maturity levels based on color and texture features. They extracted features, applied dimensionality reduction through Principal Component Analysis (PCA), and classified ripeness through an artificial neural network. They accurately distinguished fruit as raw, ripe, and half-ripe classes at 98.3% level, exhibiting great potential for visual features coupled with machine learning methodologies. It proves that it is possible to retrain the proposed model to recognize a variety of fruits other than oil palm.
Using apples, [3] constructed a smart video-based system integrating machine vision with physicochemical measurements to predict fruit ripeness. They used 444 color and texture features, and they picked the most relevant through a hybrid artificial neural network and optimization algorithms. They correctly predicted firmness, acidity, and starch content, with

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

determination coefficients greater than 0.92. Their hybrid model had a classification rate of 97.86% for ripeness stages under real orchard conditions, showing that non-invasive ripeness detection through multispectral imaging and AI is a feasible approach. [4] also performed an experiment on peaches to classify ripeness stages based on image texture parameters in different color spaces (RGB, LAB, XYZ). Classic machine learning models including Bayesian networks and Random Forests reached accuracy levels up to 100% in categorizing unique ripeness phases. Their findings identified robust correlations between ripeness and attributes such as fruit hardness, ethylene generation, and soluble solids content, and proved that texture-based visual information was capable of separating varied degrees of maturity. Venturing into sensor-based technology, [5] presented a new scent-based freshness detection system on the basis of colorimetric sensing combinatorics and deep convolutional neural network (DenseNet). Through the monitoring of gas emission from mango, peach, and banana at ripening stages, their system provided scent fingerprints identified by DCNNs. The model attained 97.39% accuracy on validation and 82.2% on actual test data, offering a lowcost, portable, and highly accurate non-destructive ripeness prediction solution. In another method based on fuzzy logic and hybrid AI, [6] suggested a pre-harvest ripeness estimation system using the Adaptive Neuro-Fuzzy Inference System (ANFIS). Their model utilized red-green color ratios from fruit images to classify six ripeness stages. Experimental results showed that their system performed better than traditional classifiers like SVM, KNN, and decision trees in terms of accuracy, sensitivity, and F-measure, further establishing the strength of fuzzy models in managing gradual ripeness transitions.

2.2 Ripeness Detection Using Hyperspectral and Multispectral Imaging

Hyperspectral and multispectral imaging have proven to be strong tools for non-destructive detection of ripeness in fruit [7]. They allow detection of the subtle biochemical and structural differences in fruits by means of reflectance at precise wavelengths. Multispectral imaging has been recently used to differentiate levels of ripeness in oil palm fruits by choosing the best wavelengths from hyperspectral datasets through the use of principal component analysis [8]. This has made possible accurate classification on the basis of internal developments instead of outer appearance alone. In the same vein, hyperspectral imaging has been employed to identify skin defects and rot in oranges. Through the use of segmented PCA and choosing characteristic wavelength bands in the visible and near-infrared regions, researchers attained more than 97% accuracy in classifying normal, defective, and rotten oranges [9]. In yellow peaches, hyperspectral imaging coupled with machine learning algorithms like XG Boost and SVM was used to estimate the storage time following bruising with an accuracy of up to 95%. Such a method assists in determining the freshness of fruits that look visually good. The creation of low-cost multispectral imaging systems has also facilitated effective fruit grading and classification, transcending the inefficiencies of manual inspection. These researches emphasize how imaging outside the visible spectrum increases the detection of internal ripeness cues, leading to feasible, high-throughput sorting systems [10].

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

2.3 Ripeness Detection Using Gas Sensors (Ethylene Monitoring)

Ethylene gas, which is a natural plant hormone, is an important signal for the ripening and spoilage of fruits [11]. A number of recent reports have investigated gas sensor-based methods for the real-time monitoring of ethylene. Graphene oxide sensors were synthesized to determine alterations in the concentration of ethylene being released by different fruits [12]. These sensors were found to possess high selectivity and linearity, with oranges registering the highest sensitivity. Their performance qualifies them for use in electronic nose devices in detecting ripeness. There are also advanced ethylene biosensors that have been created to mitigate large-scale spoilage issues within the supply chain. Research has investigated different chemical sensing technologies such as metal-oxide semiconductor sensors and PtO_2-decorated SnO_2 architectures that are highly sensitive at low detection levels. The sensors have been employed to detect ripeness and spoilage in fruits such as bananas, mangoes, apples, and even vegetables stored at room temperature [13]. Timedomain response and repeatability testing validated their usability for long-term ethylene detection. Together, these developments showcase the promise of ethylene sensors as dependable means of non-invasive, gas-based ripeness measurement. Incorporating them into smart packaging or storage rooms has the potential to minimize waste and enhance postharvest handling [14].

2.4 Ripeness Detection Using Deep Learning Techniques

Deep learning, and more specifically convolutional neural networks (CNNs), has been a prevalent method for detecting fruit ripeness using images [15]. CNNs are able to learn sophisticated patterns from high-volume datasets without requiring hand-extracted features [16]. Various researches have been able to utilize CNNs and architectures like AlexNet, ResNet, and MobileNet to classify stages of fruit maturity in bananas, mulberries, and oil palm fruits. These models attained high accuracies in classification—usually in excess of 95%—over several ripeness classes such as ripe, unripe, overripe, and rotten. Another study suggested an adaptive neuro-fuzzy inference system (ANFIS) for the prediction of fruit ripeness using red-green color difference, which performed better than conventional models like SVM and decision trees [17]. Another deep learning system utilized MobileNet V2 for both classification of fruit types and ripeness assessment, with near-perfect accuracy of 100% in ripeness evaluation. Transfer learning methodologies have also been utilized to fine-tune CNNs for effective and accurate classification, particularly when dealing with small datasets. For mulberries, ResNet-18 and AlexNet presented more than 98% accuracy and fast processing time, which indicates their capability for real-time deployment in fruit sorting lines. These studies reaffirm that deep learning is capable of high accuracy, scalability, and flexibility among various fruits and conditions. But robust performance is still dependent on high-quality data and model parameter tuning [18].

Although tremendous advances have been achieved in detecting fruit ripeness through color, texture, hyperspectral imaging, gas sensors, and deep learning methods, the majority of current research is based on image-centric or sensor-intensive systems that demand specialized

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

hardware and high computational power. Additionally, many models only concentrate on binary classification (ripe or unripe), ignoring the intermediate phases like semi-ripe, which are important for accurate harvesting and marketing decisions [19]. There are few investigations of the predictive value of organized physicochemical and categorical data—like Brix, softness, pH, harvest date, and fruit type—for predicting ripeness, though they are of interest and easily available in practical agricultural environments. Moreover, few efforts have focused on comparative performance assessment across several machine learning algorithms with such non-image data [20]. This research fills these gaps by suggesting a multi-class ripeness detection system based on tabular fruit features, providing a low-cost, explainable, and scalable substitute for vision-based solutions.

3. Methodology

This work suggests a machine learning methodology for predicting the ripeness phase of oranges based on a tabular dataset of physicochemical and categorical fruit characteristics. The process entails the preparation, preprocessing, and dealing with class imbalance of the dataset, selection of model, training, and evaluation. A pipeline of the whole process is shown in Fig.2.

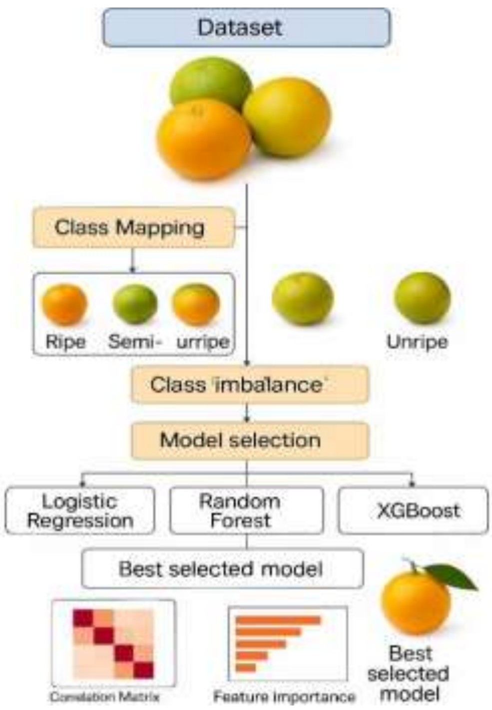


Fig.2 Workflow of the entire pipeline

3.1 Dataset Description

The dataset employed is composed of 241 orange samples with features like Size (cm), Weight (g), Brix (sweetness), pH (Acidity), Softness (1–5), Harvest Time (days), and categorical factors like Color, Variety, and Blemishes. The target variable, Ripeness, was originally

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

measured on a continuous scale from 1 to 5. To facilitate classification, these measurements were discretized into three levels: Unripe (ripeness ≤ 2.0), Semi-Ripe (2.5 ≤ ripeness ≤ 3.5), and Ripe (ripeness ≥ 4.0). The sample distribution over the ripeness categories is shown in Table 1.

Table 1: Class Distribution After Ripeness Mapping

Ripeness Class	Number of Samples
Ripe	132
Semi-Ripe	65
Unripe	44

3.2 Data Preprocessing

Preprocessing was a crucial step to get the dataset ready for successful machine learning deployment. Categorical features like Color, Variety, and Blemishes were first converted into binary indicators through one-hot encoding, and the target feature, Ripeness, was re-mapped from its numerical range (1–5) to three categorical classes: Unripe, Semi-Ripe, and Ripe. The new target labels thus generated were label encoded to make the multi-class classification possible. To gain a better insight into the distribution and discriminatory power of each numerical feature, a set of boxplots was created to represent the spread of feature values by ripeness class. The boxplots provide an indication of each feature's behavior over ranges of ripeness and assist in determining which features are likely to carry the most predictive power. The collection of boxplots, presented in Figures 3–8, captures the distribution of the following characteristics: Size (cm), Weight (g), Brix (Sweetness), pH (Acidity), Softness (1– 5), and Harvest Time (days).

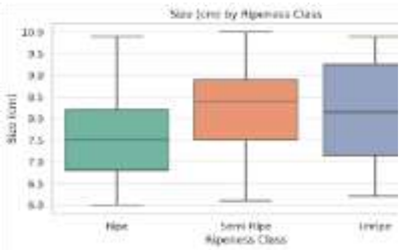


Figure 3: Size (cm) by Ripeness Class

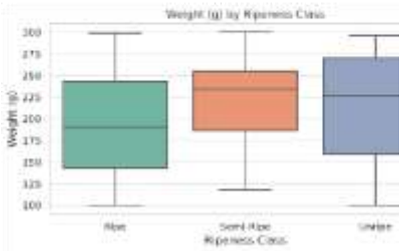


Figure 4: Weight (g) by Ripeness Class

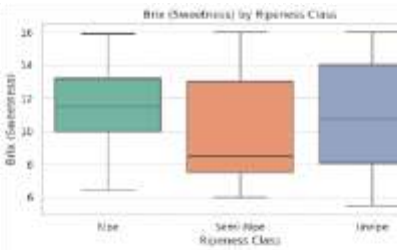


Figure 5: Brix (Sweetness) by Ripeness Class

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

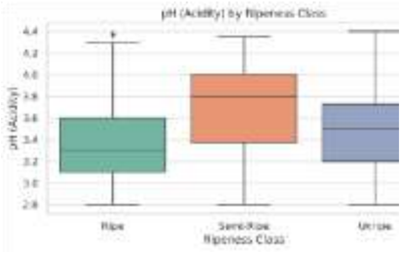


Figure 6: pH (Acidity) by Ripeness Class

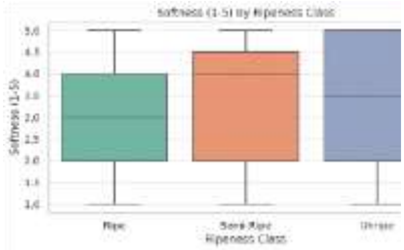


Figure 7: Softness (1–5) by Ripeness Class

Figure 8: Harvest Time (days) by Ripeness Class

For scale-sensitive machine learning models—Logistic Regression, Support Vector Machines, and Multilayer Perceptron (MLP)—all numeric features were z-score normalized. This prevented the features such as Size, Weight, and Brix from unfairly dominating the model by virtue of being in different units or scales.

3.3 Class Imbalance Handling

The initial dataset was severely imbalanced, with the majority of the samples being tagged as "Ripe." This would cause model bias, i.e., the classifier would over-predict the majority class. To overcome this, Synthetic Minority Over-sampling Technique (SMOTE) was used for the training set to generate new samples synthetically for minority classes (Unripe and SemiRipe). This approach generated a better-balanced training set, thus enhancing the classifier's capacity to identify minority classes. SMOTE was used to create new instances synthetically for minority classes by interpolating feature values between samples and their nearest neighbors. The synthetic samples were dynamically generated in such a way that the final sample size for each minority class equaled the majority class (132 instances). This led to the inclusion of 67 samples for the Semi-Ripe class and 88 for the Unripe class, resulting in a completely balanced training dataset.

3.4 Model Selection

To identify the most suitable classification model for fruit ripeness analysis, five different machine learning models were chosen and compared. The models were chosen because of their well-documented success in multi-class classification tasks as well as their applicability in previous agricultural data science research. The first model used was Logistic Regression, which was taken as a baseline given its interpretability and simplicity. Being a linear classifier, it offered a yardstick by which other more advanced models could be compared. Random Forest, as an ensemble learner built on decision trees, was utilized due to its resistance to overfitting and its power in identifying non-linear relationships. It also has the benefit of being

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

able to compute feature importance, which enhances interpretability. Furthermore, XG Boost (Extreme Gradient Boosting) was added for its high-performance boosting algorithm reputation. XG Boost is designed to be fast and accurate and has shown excellent performance on many structured data classification tasks. Support Vector Machine (SVM) was also taken into account because it can deal with high-dimensional data and is capable of identifying the best decision boundaries using kernel tricks. Lastly, the Multilayer Perceptron (MLP), a feedforward artificial neural network, was utilized to capture complicated, non-linear relationships in the data. The models were created and run under the Scikit-learn and XGBoost libraries within the Python programming framework.

3.5 Model Training and Evaluation

Following data preprocessing and class balancing with SMOTE, the dataset was split into training and test subsets in an 80:20 ratio. Stratified sampling was used to maintain the distribution of ripeness classes the same in both sets. All five models were trained on the resampled training set, and their performance was then tested on the hold-out test set. In order to fully evaluate model performance, several evaluation metrics were considered. Accuracy was computed to find the overall percentage of correctly classified instances. But since accuracy itself might not be representative of performance in multi-class problems—particularly under class imbalance—other metrics like Precision, Recall, and F1-score were also adopted to analyze class-wise performance. These measures give greater insight into the classifier's capacity to differentiate between highly similar categories, like Semi-Ripe and Ripe. In addition, a Confusion Matrix was examined to determine patterns of misclassification, providing a visual representation of the model's strengths and weaknesses in discriminating between the three classes of ripeness. Collectively, these measures provide a strong framework for assessing model performance in a multi-class classification problem.

4. Results and Discussion

This part discusses the experimental results of five machine learning models tested for orange ripeness classification into Unripe, Semi-Ripe, and Ripe classes. Performance metrics were calculated using a stratified test set and include accuracy, precision, recall, F1-score, and confusion matrix analysis. Of the models under test, Random Forest had the best overall accuracy at 75%, followed by XG Boost at 74%. The MLP and SVM models had moderate performances with accuracy measures of 66% and 65%, respectively, while Logistic Regression was the poorest at 56%. The relative model accuracies are shown in Table 2.

Table 2: Accuracy of Classifiers on Ripeness Prediction

Model	Accuracy
Logistic Regression	56%
Random Forest	75%

XGBoost	74%
Support Vector Machine (SVM)	66%
Multilayer Perceptron (MLP)	65%

4.1 Class-wise Performance Analysis

To give a more detailed insight into model behavior beyond total accuracy, class-wise performance was examined using Precision, Recall, and F1-score for every ripeness class (Unripe, Semi-Ripe, and Ripe) in all five machine learning models. The findings are represented in Figures 9–11.

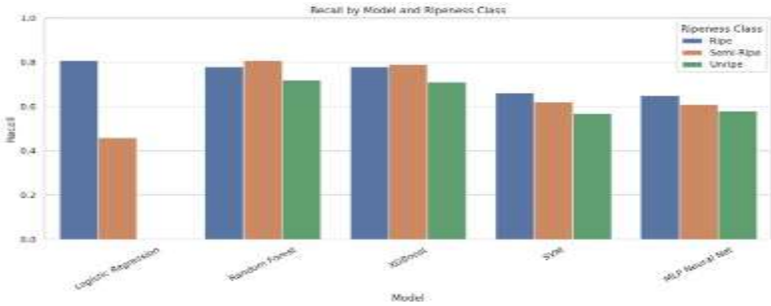
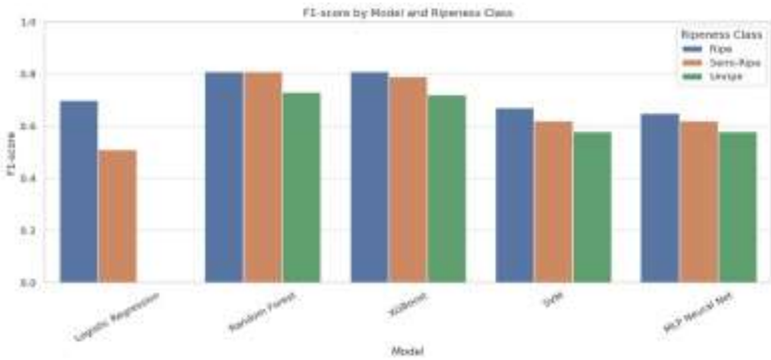


Figure 9: Recall by Model and Ripeness Class

This plot illustrates each model's performance at correctly identifying true instances of each class of ripeness. Random Forest and XGBoost both had good recall for all classes, with rates higher than 78% for Ripe and Semi-Ripe, and approximately 72% for Unripe. Logistic Regression had high recall for Ripe (81%) but did poorly for detecting Unripe, correctly classifying none. This corroborates its poor adaptability in dealing with skewed or nonlinearly separable classes. SVM and MLP models had moderate performance, with recall rates typically ranging from 57% to 66%.



International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Figure 10: F1-score by Model and Ripeness Class

F1-score, the harmonic mean of recall and precision, offers a balanced representation of model performance. Random Forest and XGBoost produced the highest F1-scores for all ripeness classes (≥ 0.79) consistently, showing excellent precision-recall balance. MLP and SVM models generated moderately accurate F1-scores (approximately 0.65–0.68 for Ripe), whereas Logistic Regression was once more the weakest with an F1-score of less than 0.52 for Semi-Ripe and zero for Unripe.

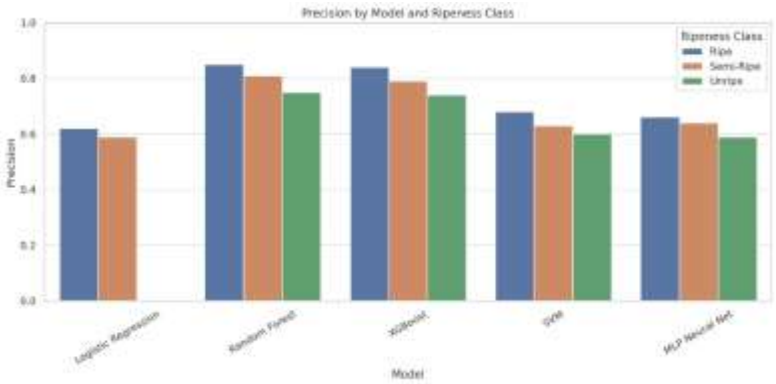


Figure 11: Precision by Model and Ripeness Class

Precision indicates the ratio of accurately predicted positives to all predicted positives. Random Forest was in the lead with precision rates of over 85% for the Ripe class and 81% for Semi-Ripe. XGBoost was close behind with similar precision for all classes. Logistic Regression, although having reasonable precision for Ripe (62%), fared badly for Semi-Ripe (59%) and failed utterly for Unripe. SVM and MLP had comparatively stable precision (~60%) for all three classes, albeit lower than tree-based models.

These performance measures highlight that Random Forest and XGBoost not only registered the highest overall accuracy but also provided balanced and stable predictions for all categories of ripeness. Logistic Regression, though interpretable, exhibited poor minority class generalization for classes like Unripe. This class-wise performance confirmation justifies the choice of ensemble-based classifiers as being ideal for multi-class fruit ripeness prediction, particularly when there are overlapping class distributions and class imbalance.

4.2 Confusion Matrix – Random Forest Classifier

The confusion matrix shows good classification performance in all categories of ripeness, with high accuracy in detecting Ripe (81%) and Semi-Ripe (78%) samples. However, the model showed moderate confusion among adjacent classes. The confusion matrix of the Random Forest classifier, as illustrated in Fig.12, emphasizes the class-wise prediction performance. The model correctly predicted 21 Ripe, 21 Semi-Ripe, and 18 Unripe samples. But there was some misclassifying between next-adjacent ripeness classes, as anticipated in biological processes where changes occur gradually and not discretely.

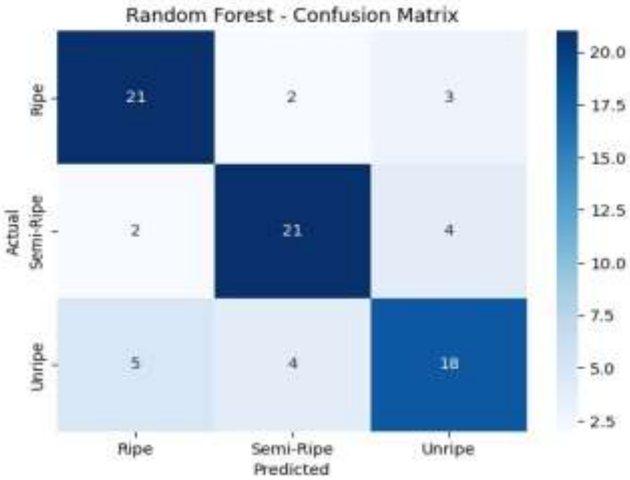


Fig.12: Confusion matrix for Random Forest

For example, five Unripe fruits were classified as Ripe, and four were classified as SemiRipe. Similarly, four Semi-Ripe samples were classified as Unripe, as depicted in Table 3. Such misclassifications point to the inherent difficulty in sharply distinguishing continuous biological processes such as fruit ripening into hard class boundaries.

Table 3: Classification of Ripeness Prediction

	Predicted Ripe	Predicted Semi-Ripe	Predicted Unripe
Actual Ripe	21	2	3
Actual Semi-Ripe	2	21	4
Actual Unripe	5	4	18

Actual Unripe 5, 4, 18, these mistakes are caused by co-varying feature values for main variables like Brix (sweetness) and Softness, which develop incrementally over ripening and usually have equal values at stages of transition. These observations indicate the potential drawback of utilizing tabular physicochemical properties in isolation and indicate that inclusion of further modalities of data, like spectral imaging, gas composition analysis, or machine vision, might yield an increase in classification accuracy for future work. Moreover, its use in remedying class imbalance was an essential factor in promoting stable performance on all classes. As indicated in Fig. 13, the initial dominance of the Ripe class was rectified by synthetic oversampling of Unripe and Semi-Ripe samples. This correction enhanced the generalization of the model to minority classes and avoided majority class bias.

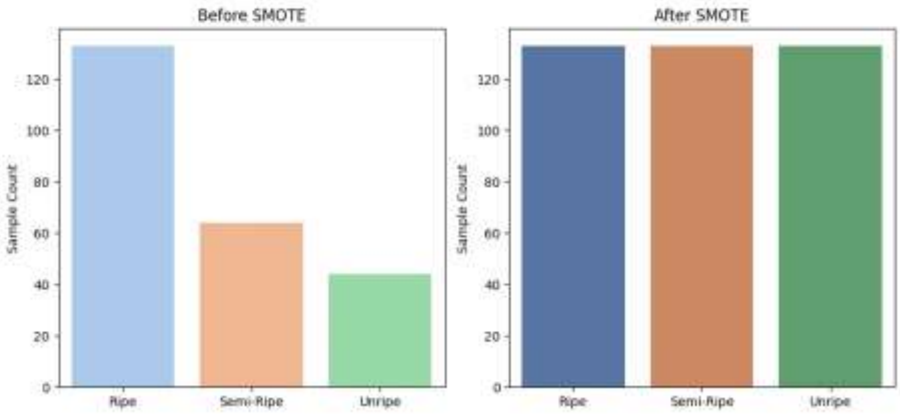


Figure 13: Class Distribution Before and After SMOTE

In order to clarify further the patterns of input features and ripeness classes, boxplot plots were examined (Figures 3–8). From boxplots, significant trends consistent with model performance and feature importance examination were revealed. Perhaps most importantly, Softness, Brix (Sweetness), and Harvest Time demonstrated clear separation between ripeness classes. For instance, Softness increased with ripeness as intuitively expected in natural fruit ripening. Brix levels—sugar content—also followed higher levels in riper fruits, whereas Harvest Time showed a linear rise, which signifies that fruits picked later tend to be ripe. These tendencies justify why these attributes ranked top among the Random Forest model's feature importance rankings. Features like Size and Weight showed overlapping ranges in classes, suggesting that they would be less successful in class separation. Likewise, pH (Acidity) also had a relatively narrow spread, diminishing its discrimination. Quality scores, while somewhat enlightening, also overlapped and probably functioned more as a validation measure than as a good predictor.

These graphical trends confirm the model's behavior and reinforce that certain features—specifically Softness and Brix—have a strong bearing on fruit ripeness predictions, while others probably add noise or redundancy. The study further highlights the prospective utility of merging other modalities in subsequent investigations, including gas sensors or imagebased features, to better track ripeness factors that are harder to detect by conventional physicochemical information.

4.3 Feature Importance Analysis – Random Forest Classifier

The feature importance analysis by means of the Random Forest model further indicated that the most significant predictors for ripeness were Softness, Brix, and Harvest Time. All these features correlate with established physiological markers in citrus fruit ripening. On the other hand, according to Fig. 14 variables such as Size, Weight, and pH showed poor predictive power, indicative of their weaker contribution to this particular classification problem.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

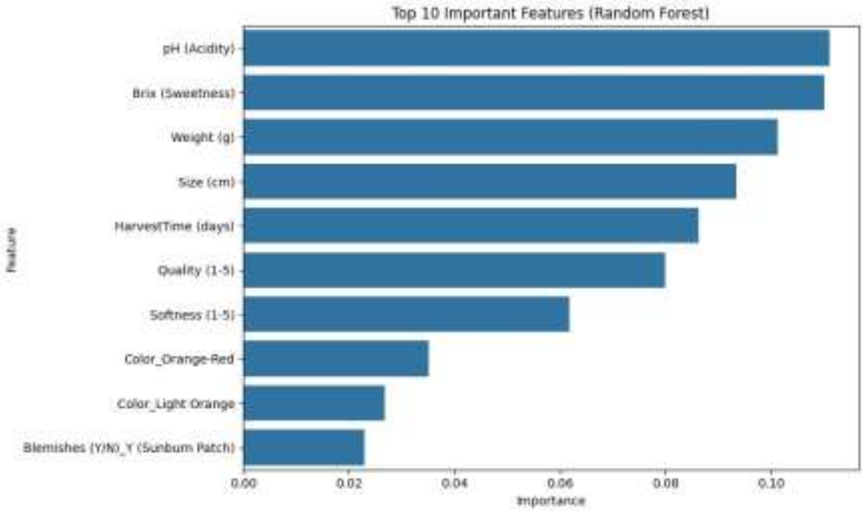


Figure 14: Feature Importance Analysis – Random Forest Classifier

In conclusion, the findings affirm the efficacy of machine learning algorithms, specifically ensemble algorithms, in making precise predictions of fruit ripeness based on structured physicochemical information. Among the classifiers tested, Random Forest exhibited the best performance, aided by stable class-wise accuracy and interpretable feature importance. The research also confirmed some of the most important predictive features—e.g., softness, sweetness (Brix), and harvest time—that are consistent with previously known ripening biological indicators. Nevertheless, remaining misclassifications among consecutive ripeness classes demonstrate the limitations of approximating continuous biological processes with discrete classes. The results suggest potential directions for future research, such as the use of multimodal sensor data and regression-based models to improve classification accuracy in real-world agricultural environments.

5. Conclusion

This work introduces a strong and interpretable machine learning-based framework for orange ripeness classification based on structured physicochemical and categorical attributes. In contrast to conventional methods that heavily depend on image or hyperspectral data, the approach in this work employs available and inexpensive tabular attributes like softness, Brix, and harvest time. A thorough analysis of five machine learning models demonstrated Random Forest as superior, with an overall test set accuracy of 75%, and robust prediction ability in all three ripeness categories—Unripe, Semi-Ripe, and Ripe. SMOTE use successfully class-distributed and increased sensitivity to the underrepresented categories. Feature importance analysis also assured the biological significance of top features, upholding the interpretability and usefulness of the model. Though certain misclassifications between sequential stages of ripeness did occur based on overlap of features, overall results establish the potentiality of using non-image data in reliable fruit classification. The proposed approach presents an expandable and adjustable pipeline that may be implemented to predict the ripeness for different fruits as well as deployed to various post-harvest technologies, smartphone apps, or quality inspection applications. Future studies can investigate the combination of this organized data with sensory modalities such as image or gas-based features to further boost model robustness and decision accuracy.

6. Reference

[1] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," Artificial Intelligence in Agriculture, vol. 7, pp. 44–57, Mar. 2023, doi: 10.1016/j.aiia.2023.02.004.

[2] Septiarini, A. Sunyoto, H. Hamdani, A. A. Kasim, F. Utaminingrum, and H. R. Hatta, "Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features," Scientia Horticulturae, vol. 286, p. 110245, May 2021, doi: 10.1016/j.scienta.2021.110245.

[3] S. Sabzi, M. Nadimi, Y. Abbaspour-Gilandeh, and J. Paliwal, "Non-Destructive estimation of physicochemical properties and detection of ripeness level of apples using machine vision," International Journal of Fruit Science, vol. 22, no. 1, pp. 628– 645, Jul. 2022, doi: 10.1080/15538362.2022.2092580.

[4] E. Ropelewska and K. P. Rutkowski, "The classification of peaches at different ripening stages using machine learning models based on texture parameters of flesh images," Agriculture, vol. 13, no. 2, p. 498, Feb. 2023, doi: 10.3390/agriculture13020498.

[5] M. Zhao, Z. You, H. Chen, X. Wang, Y. Ying, and Y. Wang, "Integrated fruit Ripeness Assessment System based on an artificial olfactory sensor and deep learning," Foods, vol. 13, no. 5, p. 793, Mar. 2024, doi: 10.3390/foods13050793.

[6] S. Kaur, S. Randhawa, and A. Malhi, "An efficient ANFIS based pre-harvest ripeness estimation technique for fruits," Multimedia Tools and Applications, vol. 80, no. 13, pp. 19459–19489, Feb. 2021, doi: 10.1007/s11042-021-10741-2.

[7] M. Shiddiq, H. Herman, D. S. Arief, E. Fitra, I. R. Husein, and S. A. Ningsih, "Wavelength selection of multispectral imaging for oil palm fresh fruit ripeness classification," Applied Optics, vol. 61, no. 17, p. 5289, May 2022, doi: 10.1364/ao.450384.

[8] H. Zhang, Y. Chen, X. Liu, Y. Huang, B. Zhan, and W. Luo, "Identification of common skin defects and classification of early decayed citrus using hyperspectral imaging technique," Food Analytical Methods, vol. 14, no. 6, pp. 1176–1193, Jan. 2021, doi: 10.1007/s12161-020-01960-8.

[9] B. Li et al., "Detection storage time of mild bruise's yellow peaches using the combined hyperspectral imaging and machine learning method," Journal of Analytical Science & Technology, vol. 13, no. 1, Jul. 2022, doi: 10.1186/s40543-022-00334-5.

[10] S. Gaikwad, "Literature Review on Multi-Spectral Imaging for Fruits and Vegetable," SSRN Electronic Journal, Jan. 2021, doi: 10.2139/ssrn.3905180.

[11] M. Mallick, D. Basu, S. M. Hossain, and J. Das, "Ethylene sensor based on graphene oxide for fruit ripeness sensing application," Applied Physics A, vol. 129, no. 2, Jan.

2023, doi: 10.1007/s00339-023-06413-7.

[12] S. Kim et al., "Highly sensitive and stable in situ acetylene detection in transformer oil using Polyimide-Embedded carbon nanotubes," Advanced Materials, Nov. 2024, doi: 10.1002/adma.202410179.

[13] X. Zhang et al., "Response Characteristics Study of Ethylene Sensor for Fruit Ripening under Temperature Control," Sensors, vol. 23, no. 11, p. 5203, May 2023, doi: 10.3390/s23115203.

[14] B. O. Olorunfemi, N. I. Nwulu, O. A. Adebo, and K. A. Kavadias, "Advancements in machine visions for fruit sorting and grading: A bibliometric analysis, systematic review, and future research directions," Journal of Agriculture and Food Research, vol. 16, p. 101154, Apr. 2024, doi: 10.1016/j.jafr.2024.101154.

[15] N. Aherwadi, U. Mittal, J. Singla, N. Z. Jhanjhi, A. Yassine, and M. S. Hossain, "Prediction of fruit maturity, quality, and its life using deep learning algorithms," Electronics, vol. 11, no. 24, p. 4100, Dec. 2022, doi: 10.3390/electronics11244100.

[16] E. Tapia-Mendez, I. A. Cruz-Albarran, S. Tovar-Arriaga, and L. A. MoralesHernandez, "Deep Learning-Based Method for Classification and Ripeness Assessment of fruits and vegetables," Applied Sciences, vol. 13, no. 22, p. 12504, Nov. 2023, doi: 10.3390/app132212504.

[17] S.-H. M. Ashtiani, S. Javanmardi, M. Jahanbanifard, A. Martynenko, and F. J. Verbeek, "Detection of mulberry ripeness stages using deep learning models," IEEE Access, vol. 9, pp. 100380–100394, Jan. 2021, doi: 10.1109/access.2021.3096550.

[18] C. C. Ukwuoma, Q. Zhiguang, M. B. B. Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, "Recent advancements in fruit detection and classification using deep learning techniques," Mathematical Problems in Engineering, vol. 2022, pp. 1–29, Jan. 2022, doi: 10.1155/2022/9210947.

[19] Y. Fu, M. Nguyen, and W. Q. Yan, "Grading methods for fruit freshness based on deep learning," SN Computer Science, vol. 3, no. 4, Apr. 2022, doi: 10.1007/s42979022-01152-7.

[20] S. Ashari, G. J. Yanris, and I. Purnama, "Oil Palm Fruit Ripeness Detection using Deep Learning," SinkrOn, vol. 7, no. 2, pp. 649–656, May 2022, doi: 10.33395/sinkron.v7i2.11420.