Comunicata Scientiae
Horticultural Journal

ORIGINAL ARTICLE
published: 30 June 2026
https://doi.org/10.14295/CS.v17.4392

Model for classifying the maturity of Biloxi blueberries using UVVIS-NIR hyperspectral images and chemometric modeling.

Jimy Oblitas[1], Edin Alva-Plasencia[2], Andre Rodríguez[2], Isabel del Rocío Pantoja Alcántara[2]

[1]Universidad Privada del Norte, Cajamarca - Peru
[2]Universidad Nacional de Cajamarca, Cajamarca - Peru
*Corresponding author, e-mail: jimy.oblitas@upn.edu.pe

Abstract

The main characteristic to define quality conditions and marketing destinations for blueberries is their ripeness, which is directly related to the internal conditions of the fruit. This creates the need for a technique that allows for rapid and individual measurement of this product. The objective of this research was to use hyperspectral images to predict the ripeness of Viloxy blueberries using chemometric analysis. For this purpose, an HSI-VNIR system was used, which works by reflectance and consists of a hyperspectral camera with a range of 400-1100 nm. Nine prediction models based on machine learning systems were developed, with the SVM model showing the best average performance, reaching an accuracy of 0.97 and an F1_macro ≈ 0.97, followed by XGBoost (accuracy ≈ 0.96, F1_macro ≈ 0.96) and Random Forest (accuracy ≈ 0.95, F1_macro ≈ 0.95). These values demonstrate that the non-destructive technique of hyperspectral imaging can be implemented for the classification and prediction of the ripeness of blueberries, thereby generating a useful tool for deciding on the marketing destinations of these fruits.

Keywords: blueberry, chemometrics, hyperspectral images, ripeness

Introduction

The berry sector in Peru is an important agent of development in rural areas and, as such, it needs to be promoted, encouraged, and its capacities enhanced in order to contribute increase its contribution to improving conditions for agricultural producers. Blueberries in Peru have generated export opportunities due to the existence of opposite seasons between Peru and the United States (Borda et al., 2018).

Blueberries (Vaccinium ssp.) belong to the Ericaceae family, which contains around 450 different species (Michalska & Łysiak, 2015). Vaccinium plants differ in shape and range from epiphytes to vines or berry trees, with most being terrestrial shrubs (Rashidinejad, 2020).

Bruising is one of the most common types of damage to blueberry tissue, and it is difficult to monitor externally due to the dark color of the fruit's skin; however, its effect severely compromises both the post-storage life and shelf life of the fruit (Xu et al., 2015). In addition to the damage that impacts to the fruit can cause (bruising), mechanically harvested blueberries are more susceptible to deterioration caused by improper removal of the pedicel, which can lead to greater weight loss and can be a vehicle for fungal diseases (Moggia et al., 2017). The ripeness of the berries is essential for good post-harvest handling, and the selection of blueberries could be based on uniformity of color and size (Brondino et al., 2022)

Another aspect of blueberry quality is related to quantitative variations in sugars, organic acids, and phenolic compounds in blueberry fruit, which are mainly associated with genetic background; particularly in their interaction with environmental factors, ripeness, cultivation techniques, and post-harvest handling (Fotirić Akšić et al., 2019), which greatly complicates estimates of export quality.

Among the inspection methods currently in use

Received: 22 February 2026
Accepted: 24 February 2026

are: visual inspection, despite being laborious, tedious, and subjective (De-la-Torre et al., 2020); chemical analytical methods, whose disadvantage lies in being destructive, time-consuming, unable to handle large numbers of samples, and costly (Shicheng et al., 2021)

Determining fruit quality using hyperspectral imaging is an objective, reliable, fast, and economical technology that offers an automated alternative to visual inspection and has been successfully implemented in various food and agricultural classification systems (Wieme et al., 2022). Among the most notable applications are those carried out on blueberries (Yang et al., 2014), coffee (Castro et al., 2018), potatoes (López-Maestresalas et al., 2016), among others.

Hyperspectral imaging technology is a non-destructive detection technology that can provide both spectral and image information. A hyperspectral image consists of a series of monochromatic images. Each spatial position of the sample can be studied from hundreds of adjacent bands. Each pixel in the image contains the spectrum of that specific location (Castro et al., 2018). In recent years, national and foreign researchers have used hyperspectral imaging technology to perform non-destructive detection of fruit diseases, including blueberries (Hu et al., 2016; Shicheng et al., 2021). The resulting spectrum acts as a fingerprint, which can be used to estimate the composition of that pixel (Wieme et al., 2022).

Along with this, it is necessary to be able to apply machine learning techniques to the spectra obtained from hyperspectral images (Wei et al., 2026). This type of analysis has shown considerable potential as it improves the accuracy, reliability, and efficiency of detecting quality problems. However, these techniques are still subject to inherent limitations (Wang et al., 2025).

For this reason, this research sought to determine whether there is a statistically significant relationship between the effectiveness of classification in determining the quality based on texture, color, and integrity of blueberries (Vaccinium myrtillus) produced in Cajamarca using multispectral images and artificial intelligence.

Materials and Methods

Sample collection

750 blueberries (Vaccinium corymbosum L.) of the Biloxi variety were collected from the same experimental plot in the Cajamarca region of Peru. Varietal homogeneity and control of growing conditions were ensured. The samples were grouped into three stages of physiological maturity for spectral analysis.

A Pika L hyperspectral imaging system (Resonon Inc.) equipped with a sensor sensitive in the range of 390–1050 nm was used. Before de acquisition of the imagines, the resolution patron, the black and the white patron was used to calibrate the HSI machine (Rodríguez-León et al., 2026). The acquisition was carried out under standardized halogen lighting (25 °C and 50% RH), with a spectral resolution of 2.1 nm per band. In order to ensure good resolution, the berries were arranged on uniform trays.

Ripeness/Maturity index

For this study, three levels of ripeness were determined, ranging from green to ripe. The ripeness index was determined by the ratio between the total soluble solids content and the titratable acidity.

The solids content was measured in clear blueberry juice, which was obtained by centrifuging and filtering it and then measuring it with an Atago digital refractometer. The values obtained are in degrees Brix. Titratable acidity was obtained from the same clarified juice by titrating it with 0.1 N NaOH in 10 mL of juice, topped up with deionized water. The results were expressed in meq/L.

Firmness

Texture was determined in a sample of 50 berries for each stage of ripeness, selecting those without mechanical damage that were representative of each ripeness class. The texturometer used was a Brookfield CT3 texture analyzer (model CT3-100, Brookfield Engineering Laboratories, MA). Measurements were taken in the equatorial zone of the fruit under the same thermal conditions of 20°C.

Spectral processing and data pre-processing

The data processing and analysis process followed the outline shown in **Figure 1**.

The spectra were extracted in raw form and their average values per sample were organized in a 750 × 116 column matrix Savitzky-Golay smoothing (window = 11, polynomial = 2) was applied to reduce instrumental noise and enhance characteristic bands. Subsequently, SNV (Standard Normal Variate) normalization was used to correct for optical dispersion effects and improve comparability between samples.

The output value of the samples was determined according to the degree of ripeness, obtaining: E1 = green or unripe, E2 = intermediate or ripe, and E3 = blue or overripe (Rivera et al., 2022). The database was divided into an 80% training/20% validation ratio, which facilitated the balanced distribution of classes. Principal component analysis (PCA) was used to reduce dimensionality and

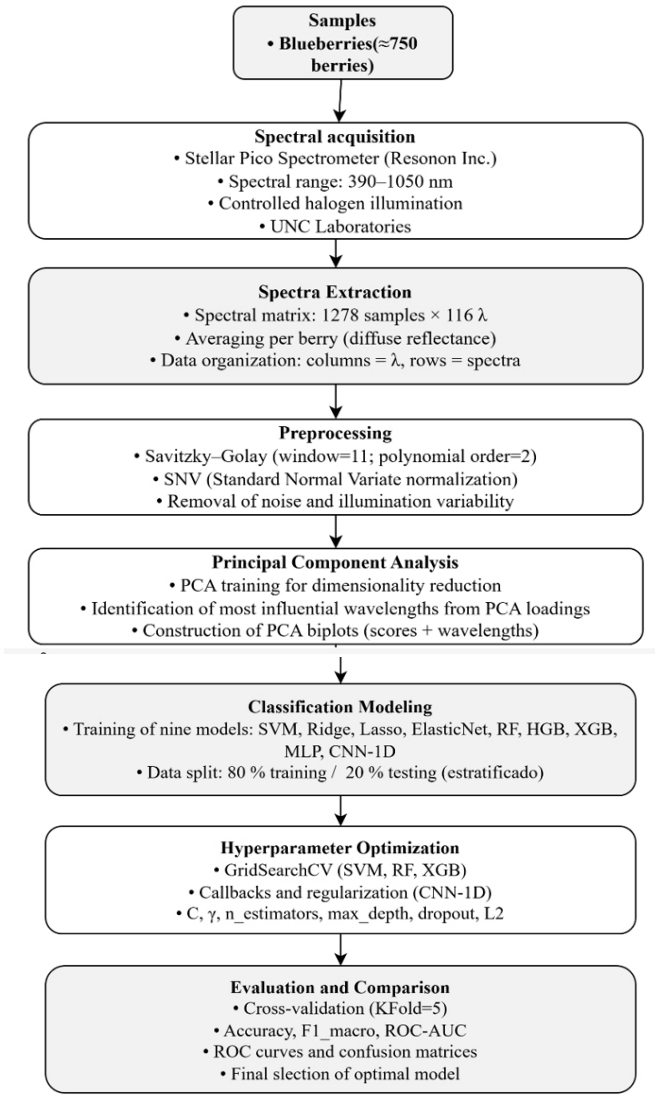


Figure 1. Working methodology

visualize spectral trends. In addition, relevant wavelengths that could be associated with changes in pigments and phenolic compounds were identified.

Training and comparative analysis of models

Nine algorithms were tested for supervised modeling: Support Vector Machine (SVM), Ridge Classifier, Lasso Classifier, Elastic Net Classifier, Random Forest (RF), Histogram Gradient Boosting (HGB), XGBoost (XGB), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN-1D).

The best-performing models were adjusted using grid search (GridSearchCV), optimizing hyperparameters such as C and gamma (SVM), n_estimators and max_ depth (RF), learning_rate and subsample (XGB), and dropout, batch_size, and L2 regularization (CNN-1D).

Finally, K-Fold = 5 cross-validation and accuracy, precision, recall, and F1_macro metrics were applied to evaluate performance. See the following equations:

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

The F1 score is first computed independently for each class using a one-vs-rest approach:

$$F1_K = 2 \cdot \frac{\text{Precision}_k \cdot \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}$$

The macro-averaged F1 score is then calculated as the unweighted mean across all classes:

$$F1_{macro} = \frac{1}{K}\sum_{k=1}^{K} F1_k$$

As part of the final comparison, confusion matrices and multi-category ROC curves were included. These visual indicators facilitated the determination of the predictive robustness of each model in relation to fruit ripeness stages.

Results and Discussion

Blueberry ripening characteristics

Figure 2 shows the evolution of the study parameters, which are the maturity index (SST/TA) and texture at the levels of the blueberry maturity under study.

When performing a statistical analysis, we observed that the maturity levels in the three stages are statistically different ($p<0.05$) from the first three levels. This has been reported in studies that observe changes in acidity and sugar composition, among others, as the blueberry ripening process progresses (Lin et al., 2020). This is due to the accumulation of sugars in the last stage of ripening associated with the high consumption of acids

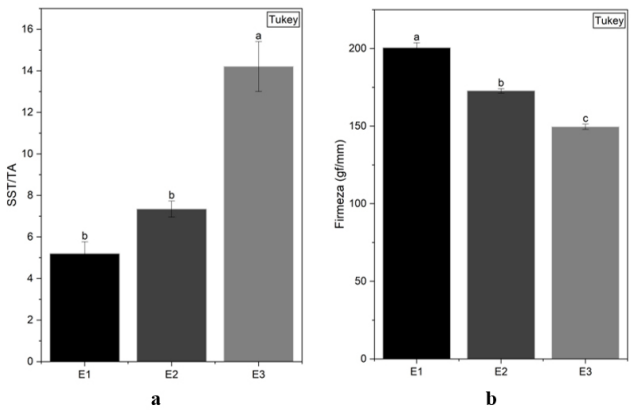


Figure 2. Analysis of maturity index (a) and firmness (b) in the different stages of maturity analyzed.

related to the active respiratory metabolism of unripe fruit, which translates into a gain in sugars and a loss of acidity in blueberries. Similar trends were reported in blueberries in previous research (Rivera et al., 2022; Shi et al., 2023).

Spectral preprocessing

The spectra were obtained from fresh Vaccinium corymbosum L. fruits at different stages of ripeness, in the range of 399 to 997 nm. A Savitzky–Golay filter (window = 9 points, order polynomial = 2) with first-order smoothing was applied, lagging instrumental noise and preserving absorption characteristics in critical regions of the spectrum (**Figure 3**).

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) reduced the dimensionality of the 112 spectra to two principal components that together explained 89.0% of the total variance (PC1 = 77.2%, PC2 = 11.8%). **Figure 4** shows the biplot (PCA + relevant lengths), highlighting how the three maturity states (E1, E2, and E3) are grouped in distinct regions of the graphic space. Unripe fruits (E1) appear mostly at high positive values of PC1, intermediate fruits (E2) are distributed closer to the center, and ripe fruits (E3) are concentrated toward negative values of PC1.

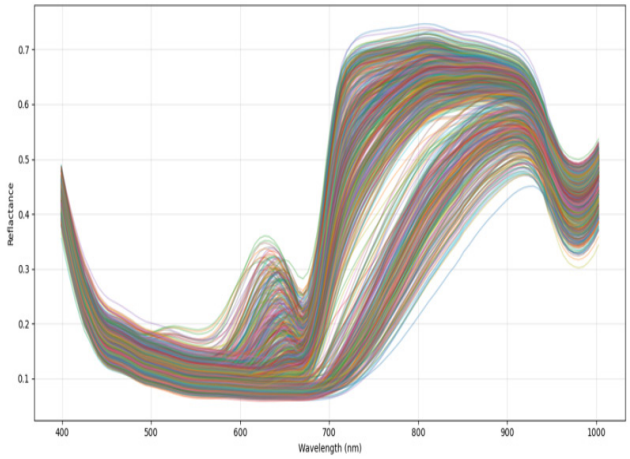


Figure 3. Absorbance spectra smoothed using a Savitzky–Golay filter

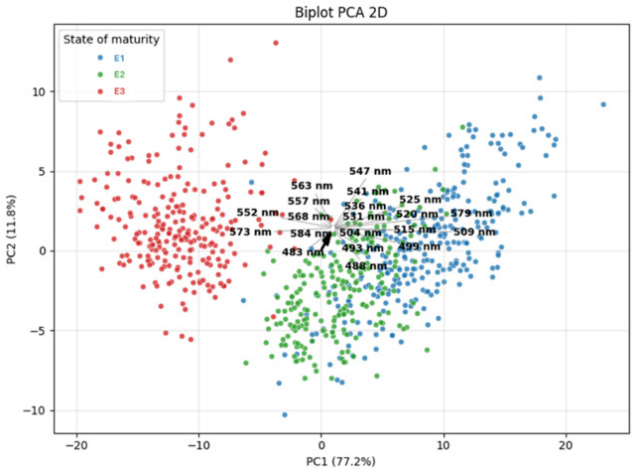


Figure 4. 2D PCA biplot and the most influential wavelengths

The observed trend is explained by the biochemical and structural changes that occur during blueberry ripening. At stage E1, the high presence of chlorophylls, fruit firmness, and low sugar content result in a characteristic spectral pattern that shows higher reflectance in the visible region and lower absorption in the near NIR (Rivera et al., 2022).

Furthermore, as the fruit ripens (E1 ↦ E3), there is an increase in the content of soluble sugars, anthocyanins, and other phenols. At the same time, chlorophyll decreases and the cell matrix is altered (Zapien-Macias et al., 2025). In addition to the above, it can be seen that the arrows on the biplot with the most influential wavelengths are 525.5, 531.0, 536.0, 541.5, 547.0, 552.0, 557.5, 563.0, and 573.5 nm. These bands are often associated with O–H (internal fruit water) and C–H (sugars) bond vibrations and with changes in anthocyanin content. Where it stands out most, the band near ~530 nm is linked to transitions from chlorophyll and green pigments to purple and blue colors, characteristic of the presence of anthocyanins. In other words, as the fruit loses chlorophyll and accumulates anthocyanins, reflectance in that range is reduced, shifting the points toward negative values in PC1.

Training and evaluation of predictive models

The total dataset (n = 750 spectra) was divided into 80% training (n = 600) and 20% testing (n = 150) using stratified sampling by classes. From the training data, nine multiclass classification algorithms were applied to determine the predictive power of the NIR spectrum. Initial results (**Table 1**) showed that deep learning and decision tree-based models performed better than linear classifiers.

Table 1. *Accuracy indicators for each trained model*

Model	Initial Accuracy	F1_macro	ROC–AUC
Ridge	0.933	0.933	0.932
Lasso	0.933	0.933	0.932
Elastic Net	0.927	0.927	0.925
HistGradient Boosting	0.940	0.940	0.994
MLP	0.940	0.939	0.994
Random Forest	0.953	0.953	0.952
XGBoost	0.947	0.947	0.993
SVM (RBF)	0.96	0.959	0.997
CNN-1D	0.96	0.957	0.998

Based on these results, the four best models (CNN-1D, SVM, RF, and XGBoost) were selected for

hyperparameter fine-tuning.

Hyperparameter optimization

The four selected models were optimized using GridSearchCV (k=5), seeking to maximize training accuracy and stability. Details are shown in **Table 2**.

Visual comparison of confusion matrices obtained for SVM, Random Forest, XGBoost, and CNN-1D models. Each panel shows the model's ability to distinguish between the three blueberry ripeness stages (E1, E2, and E3). A high concentration of values is observed on the main diagonal, indicating accurate classification. CNN-1D showed the least confusion between adjacent classes, confirming its overall best performance in predicting maturity states using NIR spectra.

Table 2. *Summary of improvement in the discriminatory capacity of the models through hyperparameter tuning*

Model	Optimized hyperparameters	Method / Range	Final Accuracy	F1-macro
SVM (RBF)	C, gamma, shrinking	GridSearch (0.1–50, 0.001–0.01)	0.970	0.951
Random Forest	n_estimators, max_depth, min_samples_split	GridSearch (100–600)	0.953	0.951
XGBoost	n_estimators, learning_rate, max_depth, colsample_bytree	GridSearch (200–800, 0.05–0.30)	0.955	0.930
CNN-1D	Dropout, LR dinámica, L2, EarlyStopping, ReduceLROnPlateau	Progressive optimization	0.973	0.972

Hyperparameter optimization improved the stability and accuracy of the selected models. For example, in SVM-clf, varying C and gamma regulated the margin and smoothness of the RBF kernel, facilitating better separation between classes (Acc = 0.970). For the Random Forest model, adjusting n_estimators and max_depth helped balance bias and variance, maintaining high generalization (Acc = 0.953).

In XGBoost, calibrating learning_rate, max_depth, and subsample improved sensitivity without overfitting (Acc = 0.955). Finally, the multilayer neural network (CNN-1D) optimized with dropout, L2 regularization, and early stopping achieved the best performance (Acc = 0.973), due to its ability to capture nonlinear relationships and complex spectral patterns associated with fruit ripening.

Cross-validation and comparative performance

The optimized models were validated using K-Fold (k=5) in order to verify their robustness and generalization capacity. **Figure 5** shows the bar chart comparison with the final performance metrics (accuracy) for the four optimized models.

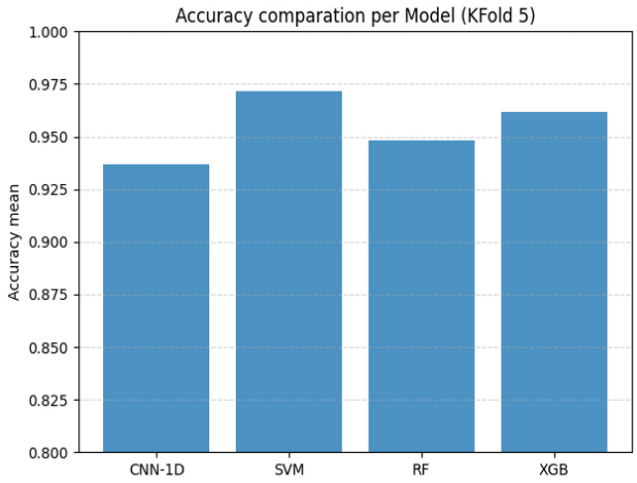


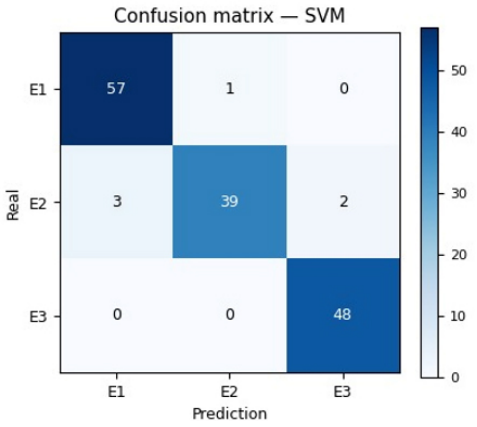
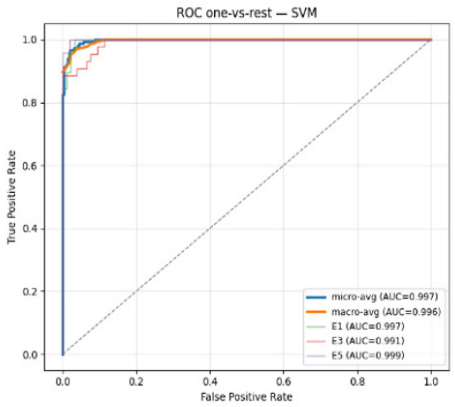
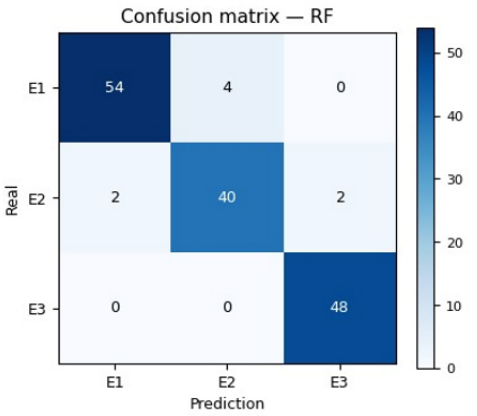
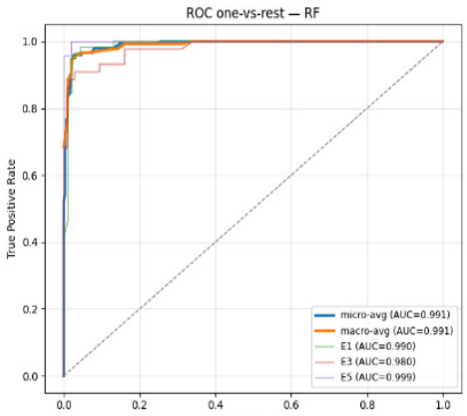
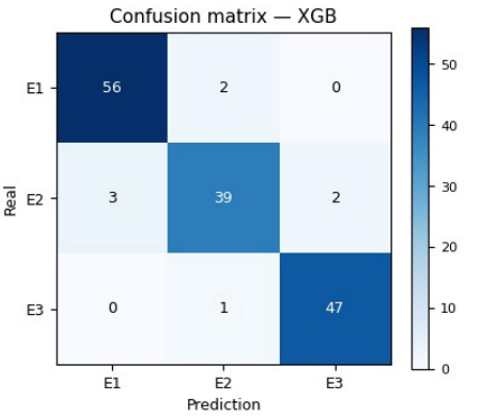
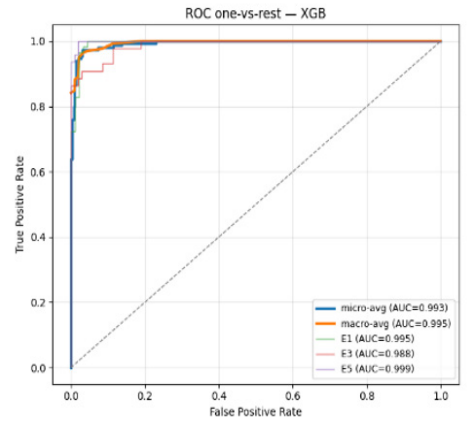
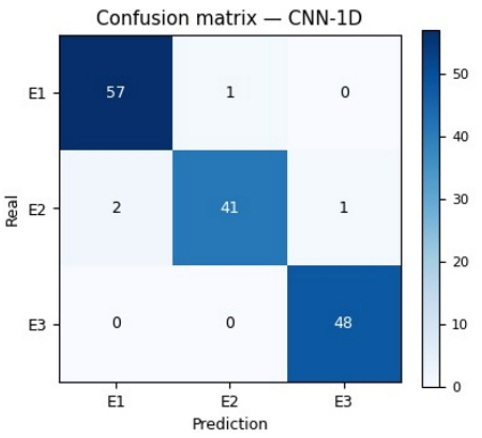
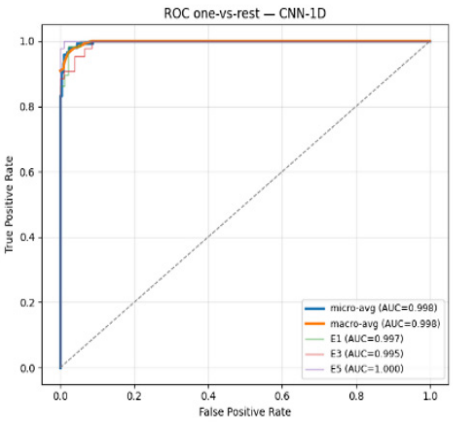
Figure 5. Comparative bar chart of average accuracy for the four optimized models.

ROC curves for the four final models. The areas under the curve (AUC > 0.95) show excellent discriminatory power for the three maturity classes. The CNN-1D network maintained a curve closer to the upper-left axis, indicating greater sensitivity and specificity. SVM and XGBoost showed very stable responses, while Random Forest showed slight variations in the intermediate probability range.

In cross-validation (KFold = 5), SVM was the model with the best average performance, reaching an accuracy of 0.97 and an F1_macro ≈ 0.97. It was followed by XGBoost (accuracy ≈ 0.96, F1_macro ≈ 0.96) and Random Forest (accuracy ≈ 0.95, F1_macro ≈ 0.95). CNN-1D obtained slightly lower values (accuracy ≈ 0.91, F1_macro ≈ 0.92), although still within a range considered high for multi-class problems with three maturity states.

Conclusions

In this study, the use of hyperspectral images based on their reflectance was evaluated, which can enable faster, more accurate, cost-effective, and non-destructive recognition of blueberry ripeness. Good predictions were obtained, with SVM being the model that performed best on average, achieving an accuracy of 0.97 and an F1_macro ≈ 0.97. This was followed by XGBoost (accuracy ≈ 0.96, F1_macro ≈ 0.96) and Random Forest (accuracy ≈ 0.95, F1_macro ≈ 0.95). CNN-1D obtained slightly lower values (accuracy ≈ 0.91, F1_macro ≈ 0.92), although still within a range considered high for multi-class problems with three maturity states, suggesting the viability of this technique to aid marketing decision-making processes. The data show potential for this to be a non-destructive measurement tool for predicting blueberry properties.

CONFUSION MATRIX	ROC CURVES
a) SVM 	a)SVM 
b)RF 	b)RF 
c)XGB 	c)XGB 
d)CNN 	d)CNN 

References

Brondino,L., Briano,R., Massaglia,S., Giuggioli,N.R. 2022. Influence of harvest method on the quality and storage of highbush blueberry. Journal of Agriculture and Food Research 10: 100415.

Castro,W., Oblitas,J., Maicelo,J., Avila-George,H. 2018. Evaluation of expert systems techniques for classifying different stages of coffee rust infection in hyperspectral images. International Journal of Computational Intelligence Systems 11: 86-100.

Cosio Borda,R., Guerra Bendezú,C., Romaní Franco,V. 2018. Opportunities for the Peruvian blueberry due to the seasonality of imports from the United States, 2014-2018. Journal of Global Management Sciences 1: 15-21.

De-la-Torre,M., Avila-George,H., Oblitas,J., Castro,W. 2020. Selection and fusion of color channels for ripeness classification of Cape gooseberry fruits. In: Mejia,J., Muñoz,M., Rocha,Á., Calvo-Manzano,J.A. (eds.) Trends and applications in software engineering*. Springer, Cham, Switzerland. p. 219-233.

Fotirić Akšić,M., Dabić Zagorac,D., Sredojević,M., Milivojević,J., Gašić,U., Meland,M., Natić,M. 2019. Chemometric characterization of strawberries and blueberries according to their phenolic profile: Combined effect of cultivar and cultivation system. Molecules 24: 4310.

Hu,M.-H., Dong,Q.-L., Liu,B.-L., Opara,U.L. 2016. Prediction of mechanical properties of blueberry using hyperspectral interactance imaging. Postharvest Biology and Technology 115: 122-131.

Lin,Y., Huang,G., Zhang,Q., Wang,Y., Dia,V.P., Meng,X. 2020. Ripening affects the physicochemical properties, phytochemicals and antioxidant capacities of two blueberry cultivars. Postharvest Biology and Technology 162: 111097.

López-Maestresalas,A., Keresztes,J.C., Goodarzi,M., Arazuri,S., Jarén,C., Saeys,W. 2016. Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging. Food Control 70: 229-241.

Michalska,A., Łysiak,G. 2015. Bioactive compounds of blueberries: Post-harvest factors influencing the nutritional value of products. International Journal of Molecular Sciences 16: 18642-18663.

Moggia,C., Graell,J., Lara,I., González,G., Lobos,G.A. 2017. Firmness at harvest impacts postharvest fruit softening and internal browning development in mechanically damaged and non-damaged highbush blueberries (*Vaccinium corymbosum L.*). Frontiers in Plant Science 8: 535.

Qiao,S., Tian,Y., Wang,Q., Song,S., Song,P. 2021. Nondestructive detection of decayed blueberry based on information fusion of hyperspectral imaging (HSI) and low-field nuclear magnetic resonance (LF-NMR). Computers and Electronics in Agriculture 184: 106100.

Rashidinejad,A. 2020. Chapter 29—Blueberries. In: Jaiswal,A.K. (ed.) Nutritional composition and antioxidant properties of fruits and vegetables. Academic Press, Cambridge, USA. p. 467-482.

Rivera,S., Kerckhoffs,H., Sofkova-Bobcheva,S., Hutchins,D., East,A. 2022. Influence of harvest maturity and storage technology on mechanical properties of blueberries. Postharvest Biology and Technology 191: 111961.

Rodríguez-León,A., Oblitas,J., Quevedo-Olaya,J.L., Vera,W., Quispe-Santivañez,G.W., Salvador-Reyes,R. 2026. Non-destructive detection of *Elasmopalpus lignosellus* infestation in fresh asparagus using VIS–NIR hyperspectral imaging and machine learning.*Foods 15: 355.

Shi,J., Xiao,Y., Jia,C., Zhang,H., Gan,Z., Li,X., Yang,M., Yin,Y., Zhang,G., Hao,J., Wei,Y., Jia,G., Sun,A., Wang,Q. 2023. Physiological and biochemical changes during fruit maturation and ripening in highbush blueberry (*Vaccinium corymbosum* L.). Food Chemistry 410: 135299.

Wang,J., Meyer,S.T., Xu,X., Weisser,W.W., Yu,K. 2025. Drone multispectral imaging captures the effects of soil mineral nitrogen on canopy structure and nitrogen use efficiency in wheat. Computers and Electronics in Agriculture 235: 110342.

Wei,C., Zhou,L., Liang,B., Chen,J., Wang,G., Li,X. 2026. Artificial intelligence revolutionize food detection? Vision, olfaction and taste integrated with machine learning/ deep learning in food detection. Food Chemistry 499: 147377.

Wieme,J., Mollazade,K., Malounas,I., Zude-Sasse,M., Zhao,M., Gowen,A., Argyropoulos,D., Fountas,S., Van Beek,J. 2022. Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: A review. Biosystems Engineering 222: 156-176.

Xu,R., Takeda,F., Krewer,G., Li,C. 2015. Measure of mechanical impacts in commercial blueberry packing lines and potential damage to blueberry fruit. Postharvest Biology and Technology 110: 103-113.

Yang,C., Lee,W.S., Gader,P. 2014. Hyperspectral band selection for detecting different blueberry fruit maturity stages. Computers and Electronics in Agriculture 109: 23-31.

Zapien-Macias,J.M., Liu,T., Nunez,G.H. 2025. Blueberry ripening mechanism: A systematic review of physiological and molecular evidence. *Horticulture Research*12: uhaf126.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.