

Anthocyanin content prediction in frozen strawberry puree

Laura García-Curiel¹, Jesús G. Pérez-Flores^{1,2*}, Elizabeth Contreras-López², Emmanuel Pérez-Escalante², Aldahir Alberto Hernández-Hernández³

¹Área Académica de Enfermería, Instituto de Ciencias de la Salud, Universidad Autónoma del Estado de Hidalgo, Circuito Ex Hacienda La Concepción s/n, Carretera Pachuca-Actopan, San Agustín Tlaxiaca, Hidalgo, México; ²Área Académica de Química, Instituto de Ciencias Básicas e Ingeniería, Universidad Autónoma del Estado de Hidalgo, Carretera Pachuca-Tulancingo, Mineral de la Reforma, Hidalgo, México; ³Área Académica de Ingeniería Agroindustrial e Ingeniería en Alimentos, Instituto de Ciencias Agropecuarias, Universidad Autónoma del Estado de Hidalgo, Avenida Universidad Km. 1 s/n Exhacienda Aquetzalpa, Tulancingo de Bravo, Hidalgo, México

*Corresponding Author: Jesús G. Pérez-Flores, Área Académica de Química, Instituto de Ciencias Básicas e Ingeniería, Universidad Autónoma del Estado de Hidalgo, Carretera Pachuca-Tulancingo km 4.5, 42184 Mineral de la Reforma, Hidalgo, México Email: jesus_perez@uaeh.edu.mx

Received: 18 December 2022; Accepted: 8 May 2023; Published: 25 May 2023

© 2023 Codon Publications

OPEN ACCESS 

ORIGINAL ARTICLE

Abstract

Rapid color degradation during processing and storage is a limitation when using strawberry puree (SP). This work aimed to use image analysis coupled with two machine learning algorithms: ordinary least squares (OLS) and artificial neural networks (ANNs), to predict anthocyanin content (AC) in frozen SP during its storage at -18°C for 120 days. When applying the OLS regression model, unsatisfactory AC prediction values were obtained due to multicollinearity. In contrast, a good prediction of AC using ANNs model was observed by comparing AC in SP predicted by the model versus the experimentally obtained values (coefficient of determination, $R^2 = 0.977$).

Keywords: anthocyanin content, color measurement, image analysis, machine learning, strawberry puree

Introduction

Strawberry (*Fragaria × ananassa*, Duch.) fruit contains nutritional compounds, such as sugars, proteins, dietary fibers, vitamins, and minerals; bioactive compounds, such as ascorbic acid, carotenoids, flavonoids, folates; and phenolic compounds, such as anthocyanins, most of which are natural antioxidants and contribute to the high nutritional quality of the fruit (Hosseinfarahi *et al.*, 2020; Liu *et al.*, 2018; Teribia *et al.*, 2021). Strawberries are non-climacteric fruits, so they must be harvested almost at maturity to guarantee its highest quality in terms of flavor, color, and consistency (Mancini *et al.*, 2020).

Strawberry is one of the most commonly consumed berries, fresh and processed, as a concentrate, juice, or puree in the formulation of different products. In the food industry, strawberry puree (SP) is used to prepare red-colored

products with a tasty flavor, such as fruit preparations, ice creams, and smoothies. However, the use of SP in these products is limited by its rapid color degradation during processing and storage (Teribia *et al.*, 2021). This loss of red color is attributed to the degradation of anthocyanins and enzymatic and nonenzymatic browning reactions. In addition, color stability depends on many factors, such as temperature, water activity, light, oxygen, pH, and ascorbic acid (Da Silva Simão *et al.*, 2022).

Change in color during storage is a quality parameter with the most significant impact on the shelf life of fruit-based products (Buvé *et al.*, 2018), because it plays an essential role in influencing the sensory and hedonic expectations of consumers. Therefore, change in color can lead to product rejection, even on the shelves of the market (Da Silva Simão *et al.*, 2022). Besides this, studying the relationships between color and pigments is equally important.

Recently, image analysis has gained interest for its simplicity, reliability, low cost, and speed of analysis to assess food quality, in addition to the fact that it does not require reagents. Many properties can be extracted from an image, for example, color, pixels values distribution, statistical greatness, and frequency domain measures (Kato *et al.*, 2019). Color space extraction from food matrices image has been previously reported by Barbin *et al.* (2016), Ulrici *et al.* (2012), and Valous *et al.* (2009), providing a whole idea of the product instead of color measurement of a single point or a reduced area, such as the one spotted by a colorimeter (Barbin *et al.*, 2016). Hence, the implementation of a computer vision system (CVS) for predicting anthocyanin content (AC) in SP during storage constitutes a nondestructive and low-cost quality control tool that allows making decisions about rotation, applications, or processing conditions of SP when used as an ingredient in the preparation of other food products.

Computer vision, also called artificial vision, is one of the branches of artificial intelligence (AI) and is responsible for understanding in detail the visual data, similar to human optical systems, to make decisions using other branches of artificial intelligence, such as machine learning or deep learning (Xu *et al.*, 2021). Machine learning and deep learning can carry out the tasks of recognition, prediction, or classification, in which they make decisions after a trained and evaluated computational model based on a dataset (Barbin *et al.*, 2016; Santos Pereira *et al.*, 2018). Therefore, a computer vision system is based on the following stages: (1) image acquisition, (2) image segmentation, (3) image feature extraction and selection, and (4) image classification, object detection, or feature prediction using machine learning or deep learning methods (Contreras-López *et al.*, 2022; Lopes *et al.*, 2019; Oliveira *et al.*, 2021).

The basis of deep learning is artificial neural networks (ANNs). ANNs are advanced fitting and pattern recognition algorithms that allow users to extract complex relationships among nonlinear variables. As the training progresses, the neural network-based model learns the unknown dynamics of the process. This advantage makes ANNs very appealing computational tools in applications with little or incomplete understanding of the problem, although experimental measurements are readily available (Kazi *et al.*, 2021). Further, the application of multivariate statistical methods, such as multiple linear regression (MLR), to single out color parameters to correlate them with the pigment content in strawberries has been reported as well (Amoriello *et al.*, 2022; Hernanz *et al.*, 2008). Therefore, this work aimed to use image analysis to predict AC in frozen SP during its storage, verifying the effectiveness of two machine learning algorithms—ordinary least squares (OLS) and ANNs algorithms—to build models that allow predicting AC

using 3-dimensional (3D) CIELAB color space (also referred as $L^* a^* b^*$ coordinates, it covers the entire range of human color perception), to promote its use as a quality control tool by producers, scientists, and food technologists interested in a possible commercial application.

Materials and Methods

Preparation and processing of strawberry puree

Strawberries were obtained locally from la Central de Abastos de la Ciudad de Pachuca, Pachuca de Soto, Hidalgo, México. After strawberries were washed and disinfected, the calyx and leaves were removed and the fruit was chopped. Small pieces were mashed and blended for 1 min at 15,000 rpm in a domestic Osterizer blender (Oster, Mexico). The resulting puree was passed through a 500- μm stainless steel sieve to remove seeds (under atmospheric conditions). The puree prepared was pasteurized in 100-mL glass jars (55-mm diameter and 75-mm height) at 90°C for 15 min. Pasteurized purees were cooled with water at 25°C. This process assured the puree's microbial stability (Marszałek *et al.*, 2015). Samples of pasteurized purees (100 g) were manually filled, under aseptic conditions, into small polypropylene containers (180 mL) with some headspace left in containers. Finally, samples were stored at -18°C for 120 days, determining AC and color every 5 days.

Determination of AC

The AC of SP was determined in triplicate with a modified pH differential method described by Zheng *et al.* (2011). Results were expressed as milligram cyanidin-3-glucoside equivalents per liter of puree (mg L^{-1}).

Total titratable acidity (TTA)

The TTA of SP was determined using an acid–base titration method. Fruit puree (1 mL) and distilled water (50 mL) were added in an Erlenmeyer flask. Then, a few drops of phenolphthalein were added, and the sample was titrated with aqueous NaOH 0.1 mol L^{-1} to attain a pH 8.1. Total acid contents were calculated as gram of citric acid in 100 g of sample and were presented as a mean of triplicate analyses (Kafkas *et al.*, 2007).

Color measurement by image analysis

Image acquisition system

The image acquisition system used to determine color changes in SP during storage consisted of an illumination

chamber, a charge-coupled device (CCD) digital camera, and a personal computer (PC). All were constructed, configured, and calibrated according to the research conducted by Contreras-López *et al.* (2022).

Image analysis

Red (R), green (G), and blue (B) (RGB) analysis of digital images was carried out using the ImageJ software (1.53 k; <https://imagej.nih.gov/ij/index.html>), an open-source program from the National Institutes of Health and the Laboratory for Optical and Computational Sciences, to precisely confirm color changes in samples. Image processing was carried through the Plugins/Analyze/RGB Measure menu, and average values were obtained for R, G, and B. Subsequently, these values were transformed to CIELAB color space coordinates, as indicated in the research conducted by Wu and Sun (2013).

Coordinates were expressed as L^* describing lightness ($L^*=0$ for black, and 100 for white), a^* or redness for intensity in green–red ($a^*<0$ for green and >0 for red), and b^* or yellowness describing intensity in blue–yellow ($b^*<0$ for blue and >0 for yellow), representing rectangular chromaticity coordinates. Subsequently, the overall color difference (ΔE), hue angle or color angle (h^*), and chroma or color saturation (C^*) were calculated as reported in the literature (Udomkun *et al.*, 2017; Wu and Sun, 2013).

Multiple linear regression model

The OLS regression model was applied to determine how the CIELAB coordinates were related to AC. The OLS regression model is a simple machine learning algorithm and was defined as follows (Mollalo *et al.*, 2020):

$$y_i = \beta_0 + \beta x_i + \varepsilon_i, \quad (1)$$

where y_i is the dependent variable (AC), β_0 is the intercept, β is the vector of regression coefficients, x_i is the vector of selected explanatory variables (independent variables: L^* , a^* , and b^*), and ε_i is a random error term. OLS optimizes regression coefficient (β) by minimizing the sum of squared prediction errors.

Prediction performance was evaluated using R^2 . The numerical tests were performed using Python's StatsModels library (v0.13.2). A correlation matrix heatmap was carried out using Python's Seaborn library (v0.12.1) to represent visual correlations between independent and dependent variables. In this study, the OLS regression model was trained on 70% of the dataset and tested on the remaining 30% (train_test_split random state = 100 in Python's Scikit-learn [v1.1.3]).

Artificial neural networks approach

Data preprocessing

Standardization or Z-score normalization on the CIELAB color space dataset was computed using Equation 2, which subdivides data points in terms of standard deviation away from the mean value of the distribution, as follows:

$$Z = \frac{x_i - \bar{x}}{S}, \quad (2)$$

where Z is the result of normalization value, x_i is the i th data point, \bar{x} is the sample's mean value, and S is the sample's standard deviation (Prihanditya and Alamsyah, 2020).

ANNs model

A multilayer perceptron (MLP) is a feed-forward ANNs model that consists of (1) an input layer with nodes representing independent variables, (2) an output layer with nodes representing dependent variables, that is, what is being modeled, and (3) one or more hidden layers containing nodes to help capture nonlinearity in the data (Pilkington *et al.*, 2014). In these feed-forward networks, the Levenberg–Marquardt algorithm, which is an iterative algorithm, achieves error minimization. The whole data are randomly split into training and testing groups. The training set is used to train the network whereas the test set is used to evaluate the network's performance after training (Amoriello *et al.*, 2022).

The ANNs model used in this study was based on a multilayer perceptron and was developed and trained with Python programming language (v3.8.5) using Anaconda (v4.13.0), a free and open-source distribution for managing Python libraries. The development environment configuration consists of multiple packages and libraries: Keras (v2.3.1) runs on top of the machine learning platform TensorFlow (v2.9.1). Additionally, other Python packages and libraries were used, such as NumPy (v1.23.1), Scipy (v1.7.3), Matplotlib (v3.5.1), Pandas (v1.4.3), Seaborn (v0.12.1), and ScikitLearn (v1.1.1). Finally, code was developed on a Python notebook using the Visual Studio Code software (v1.70.0) as an integrated development environment (IDE). The number of artificial neurons in input and output layers was defined as a function of dependent and independent variables, respectively. In contrast, the number of hidden layers and the number of artificial neurons required in each hidden layer were determined by trial and error to minimize the deviation of predictions from experimental results.

After defining the layers, the input data were divided into a training dataset (70% of the input data) and a testing dataset (30%). The ANNs model was compiled and trained for 2000 epochs and optimized using root mean square propagation (RMSprop) as an optimization algorithm with a learning rate of 0.01, using a random seed of 0 (random_state = 0). The mean square error (MSE) was used as a network performance index (loss function), and the mean absolute error (MAE) was used as an evaluation metric. In the *validation_split* parameter, a fraction of 30% of the training data was put aside to monitor training performance.

Finally, to use ANNs model, the calculated weights must be available for later use in other applications. The strategy used here was to export trained model in a Python pickle file (.pkl) using the Python library Joblib (v0.13.2).

This study used MAE and MSE as performance parameters to compare OLS and ANNs models (Bilgili and Sahin, 2010).

Results and Discussion

Color parameters and anthocyanin content

Figure 1 shows the evolution of the physicochemical quality attributes of SP samples during frozen storage. The results show a decrease of L^* values from 33.760 on day 0 to 19.640 on day 120 (Figure 1A), meaning that the fruit developed darker color during storage (Caner *et al.*, 2008).

Hue (h^*) is an angular value representing a dominant wavelength (Athira *et al.*, 2019) so that it characterizes color modifications: 0° (or 360°) is defined for red, 90° for yellow, 180° for green, and 270° for blue color (Scalisi *et al.*, 2022). It is observed in Figures 1B and 1C that values of a^* and b^* decreased with increase in storage time; decrease in h^* values ranged from 0.641° on day 0 to 0.568° on day 105; however, a slight increase was observed after 105 days (Figure 1D). These results implied that the sample maintained its red color during storage.

Decrease in C^* values ranged from 48.388 on day 0 to 34.058 on day 120 (Figure 1E), together with decrease in a^* , which explained decrease in the redness of SP during storage. This is because chromaticity is a measure that moves from the center of the CIELAB color space system ($C^* = 0 = \text{gray}$) to the direction of puree colors ($C^* = 100$); higher values of C^* indicate higher purity or color intensity (Contreras-López *et al.*, 2022). Finally, ΔE showed total increase in color during the test period, with value of 0.336 on day 5 to 20.121 on day 120 (Figure 1F).

Generally, this change becomes more evident when $\Delta E > 5$ (Contreras-López *et al.*, 2022).

Loss of bright red color in SP stored at freezing temperatures (-18°C) could be associated with the degradation of phenolic compounds and anthocyanins, which can occur at different temperatures (Zhang *et al.*, 2019). In fact, in this work, a decrease in AC was observed during storage (Figure 1G). Similarly, it was reported that using refrigerated temperatures (4°C) during processing did not improve the color and anthocyanin stability of SP (Teribia *et al.*, 2021). In addition, a decrease in the concentration of the total phenol content was observed in SP at freezing temperatures, but after 9 months of storage (Obradović *et al.*, 2020). In addition, phenolic compounds in strawberries, such as pelargonidin, ellagic acid, p-coumaric acid, quercetin, and kaempferol derivatives, are very unstable during freezing process because of microbial enzymes and nonenzymatic oxidation (Aaby *et al.*, 2007; Oszmiański *et al.*, 2009). Therefore, the storage temperature is an important factor in extending the shelf-life of strawberries (Lv *et al.*, 2022) and derived products, such as SP.

Moreover, anthocyanins positively correlated with antioxidant activity (Zhang *et al.*, 2019). Another factor that influences the stability of phenolic compounds is vitamin C, which decreases when the storage temperature or storage time increases compared to whole strawberries. In addition, oxidation of ascorbic acid affects the loss of flavylum pigmentation in anthocyanins (Howard *et al.*, 2014; Stan *et al.*, 2016).

On the other hand, Figure 1H shows an increase in titratable acidity (gram of citric acid per each 100 g of SP), which is associated with the ripeness stage of strawberries and color changes during freezing storage (Galoburda *et al.*, 2014; Stan *et al.*, 2016).

Prediction of anthocyanin content in strawberry purees

Ordinary least squares regression model

The OLS regression model presented an R^2 of 0.928. This indicated that the model was capable of explaining the variability of 92.80% observed in the AC of SP during storage at -18°C . This was a statistical measure of how well the regression line approximated experimental data points. The adjusted R^2 reflected model complexity and was considered a more accurate measure of model performance (adjusted $R^2 = 0.917$). On the other hand, model's p -value was significant (3.84×10^{-12}); so, the coefficients were different from 0 and could predict the dependent variable (AC). Summary statistics of the OLS regression model that described the relationship between

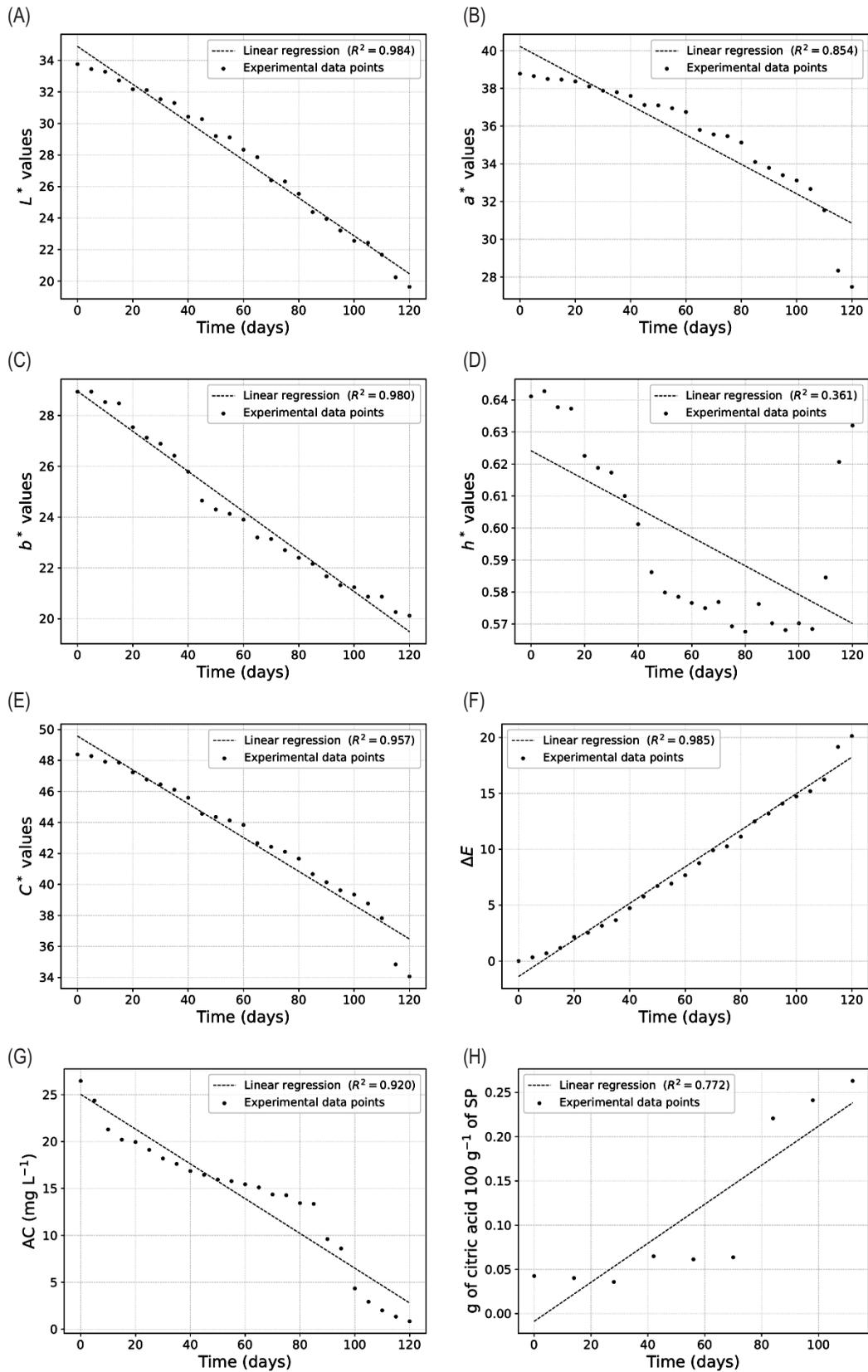


Figure 1. Scatter plots of changes in the physicochemical quality attributes of strawberry puree (SP) stored at -18°C for 120 days. (A) Lightness, (B) redness, (C) yellowness, (D) hue angle, (E) chroma or color saturation, (F) the overall color difference, (G) anthocyanin content (AC), and (H) total titratable acidity.

Table 1. Summary statistics of the OLS regression model on selected variables in modeling anthocyanin content in strawberry purees stored at -18°C .

Variable	Coefficient	Standard error	t-statistic	p-value	VIF [†]
Intercept	-41.9350	13.5577	-3.0931	0.0055	
L^*	0.9323	0.7825	1.1914	0.2468	69.7646
a^*	0.7076	0.5613	1.2605	0.2213	17.4823
b^*	0.2024	0.7647	0.2646	0.7939	28.8329

[†]Variance inflation factors (VIF) for independent variables.

the CIELAB color space coordinates and AC in SP are shown in Table 1. According to Equation 1, values of β and x_i are the following vectors: $[0.9323, 0.7076, 0.2024]^T$ and $[L^*, a^*, b^*]^T$, respectively.

Regression coefficients computed for each independent variable represent the strength and type (positive or negative) of relationship between independent and dependent variables. The statistical significance of coefficients associated with each independent variable is assessed by *t*-test. Model's coefficients with small *p* values are important. The associated variables are effective predictors (Lukawska-Matuszewska and Urbański, 2014).

Finally, one of the methods for determining the presence of multicollinearity is the Variance Inflation Factor (VIF). The VIF indicates how much the variance of a coefficient associated with explanatory variable increases because of the linear dependence between independent variables. The variables related to high VIF values are usually eliminated from the model. VIF above 5 or 10 indicates high multicollinearity between independent variables (Lukawska-Matuszewska and Urbański, 2014; Wagle *et al.*, 2017), resulting in less reliable statistical inferences. This could explain why the *p* values of coefficients of independent variables are not significant ($p > 0.05$). These results coincide with what was observed in the correlation heatmap (Figure 2A). In this correlation plot, each numerical variable represented a column, and rows showed relationship between each pair of variables. The color-coding of cells made it easy to identify visually the strength relationships (linear and nonlinear) between variables. All relationships between variables presented $R^2 > 0.880$. Generally, a Pearson correlation coefficient greater than 0.800 indicates the presence of multicollinearity (Lev *et al.*, 2022).

Finally, AC values predicted by the OLS regression model versus the true (experimental) AC values showed a linear relationship, obtaining $R^2 = 0.928$ (Figure 2B). Still, owing to the unsatisfactory results of OLS regression model, a second attempt at prediction was accomplished using ANNs to assess whether it performed better predicting AC.

ANNs model

Different ANN configurations were developed and compared to determine ANNs model with a better fitting architecture (input-hidden-output layers and artificial neural number). The ANNs model built presented the following hyperparameters and activation functions: three artificial neurons in the input layer that checked with L^* , a^* , and b^* coordinates of CIELAB color space; the input layer was fully connected to the first hidden layer that consisted of 10 artificial neurons applying the activation function as rectified linear unit (ReLU). The second hidden layer consisted of eight neurons with the same activation function as ReLU. The last layer, the output layer, received values from the second hidden layer and transformed them into output values to model the AC of SP. The activation function was used to compute the predicted output of each neuron in each layer by using inputs, weights, and biases. In the output layer, activation function was not used.

The model summary was printed to identify full-fledged parameters with training and testing. A total of 137 parameters were acquired, including trainable as well as zero non-trainable parameters. Once the model was adjusted through 2000 epochs, the loss percentage decreased more slowly. It halted after the 1000th epoch, as observed about the training history in Figure 2C, up to the 1999th epoch, the readings were: loss (MSE): 0.1805 and validation loss (validation MSE): 0.8541, the plot suggests that ANNs model has a good fit on the problem starting at 1000th epoch; when the MSE value no longer decreases, an optimal number of training cycles were reached (Rogiers *et al.*, 2012). In contrast, it is observed in Figure 2D that MAE values always descended over the epochs, leading to higher accuracy.

Conversely, the validation MAE values had a slight upward trend after 875th epoch. This suggested that ANNs model was overfitting (Palkovits, 2020), which could be attributed to the fact that few data were available. It was observed that the model stopped learning at 1500th epoch. Up to 1999th epoch, the readings were MAE: 0.2598 and validation MAE: 0.8724.

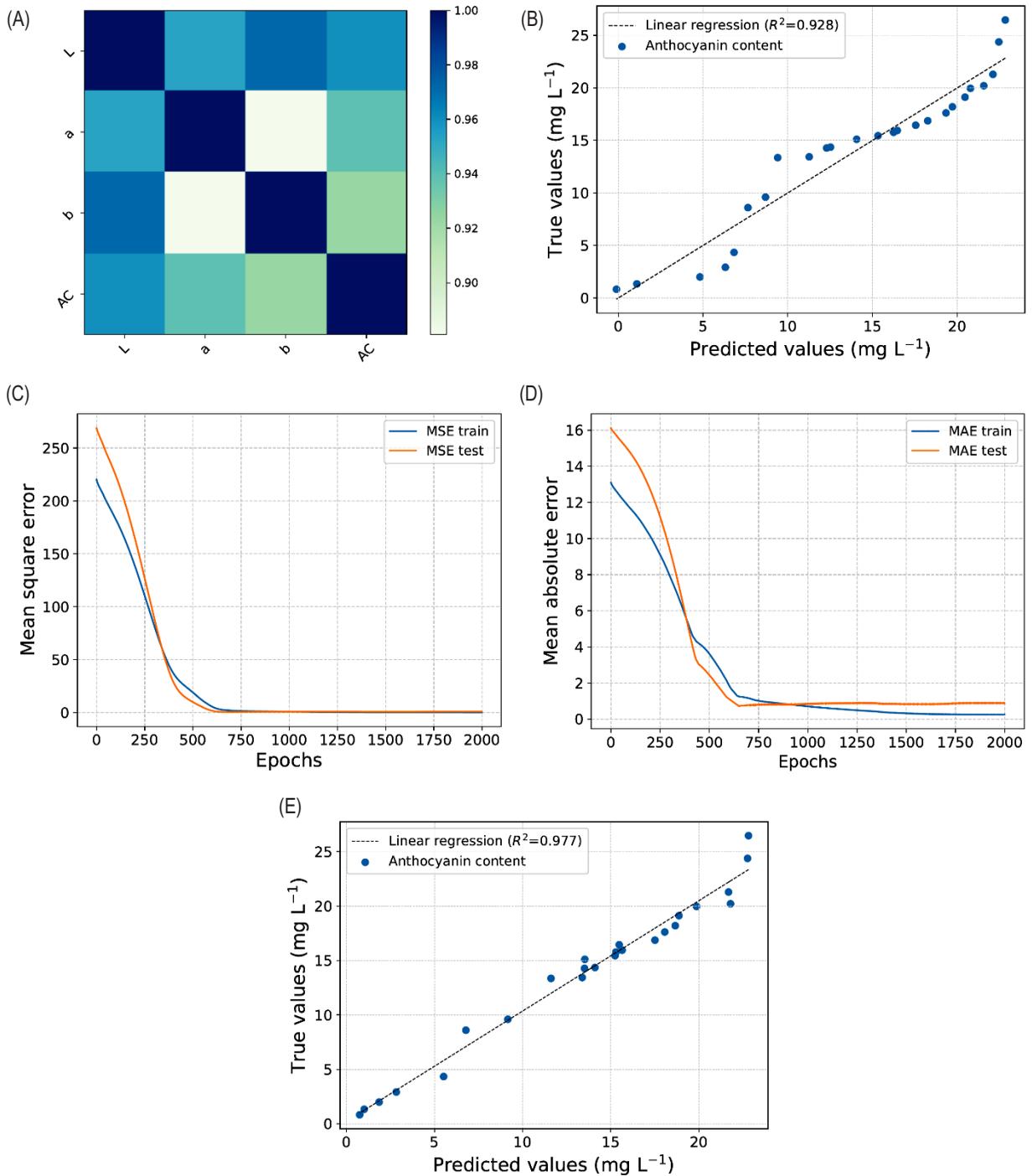


Figure 2. Results of the OLS regression model and training of ANNs model. (A) Correlation heatmap. (B) Scatter plot of the predicted vs experimental data of AC achieved by the OLS regression model. (C) Training history using the mean square error as a loss function. (D) Mean absolute error as an evaluation metric. (E) Scatter plot (predicted vs true) of the training data of AC achieved by ANNs model. The optimal regression line between estimated and measured values is shown in (B) and (D).

Finally, AC in SP, as predicted by ANNs model, was compared to the experimentally obtained values. To test the model's suitability, the predicted and actual results were plotted in Figure 2E. An R^2 of 0.977 illustrated a good agreement between two sets of results, much better than

the use of OLS regression modeling during the prediction of AC.

Table 2 compares MAE and MSE values for training and test stages of OLS and ANNs models. These results

Table 2. Performance values of OLS and ANN models.

Performance parameter	OLS regression model	ANNs model
MSE test	3.0923	0.8541
MSE train	3.7131	0.1805
MAE test	1.4188	0.8724
MAE train	1.6526	0.2598

demonstrated that the ANNs modeling approach provided more accurate results than the OLS regression model because both MAE and MSE values were small. This indicated that the best agreement between experimental and predicted AC values was obtained with ANNs model (Bilgili and Sahin, 2010; Tosun *et al.*, 2016).

The predictions were made with 30% of the training data in each case. It was observed that ANNs model better agreed with the experimental AC values (true values). However, both models could be improved by training with a larger amount of data. Therefore, selecting an adequate model for evaluating the relationship between CIELAB color space and AC was essential for predictions that are more accurate.

Finally, the use of machine learning models combined with image analysis constituted a novel, robust, cheap, and easy to implement potential tool to predict changes in the AC of SP during frozen storage as well as during storage, processing, and distribution stages. It could also be useful for analyzing other fruit-based products in which the factor with the most significant influence on shelf life is changes in visual aspects, such as color change.

Conclusion

The present study highlights that image analysis and machine learning models represent a promising, non-destructive, less time-consuming, and inexpensive tool for rapidly monitoring quality attributes of SP during frozen storage. In the future research, the best machine learning model to implement a computer vision system could be determined that fruit cultivators and food technologists could use successfully for different purposes throughout the supply chain, such as the assessment of color changes during food processing, ripening, and degradation.

Prediction of AC in SP during storage is an important quality control parameter because it is directly associated with color. This determines the shelf life of the product and its applications. It also decides processing conditions of SP in the formulation of different food products. Furthermore, titratable acidity increases in frozen strawberries with extended storage time, and

could be correlated with color change and ripeness of the fruit.

Acknowledgments

The authors thank the Sistema Nacional de Investigadores (SNI-CONACyT) and the Universidad Autónoma del Estado de Hidalgo.

Conflicts of interest

The authors declared no conflict of interest for this paper.

Author Contributions

L. García-Curiel: Writing - Original Draft, Writing - Review & Editing; J. G. Pérez-Flores: Conceptualization, Visualization, Software, Supervision; E. Contreras-López: Investigation, Resources; E. Pérez-Escalante: Formal analysis; A. A. Hernández-Hernández: Writing - Review & Editing.

References

- Aaby K., Wrolstad R.E., Ekeberg D. and Skrede G. 2007. Polyphenol composition and antioxidant activity in strawberry purees; impact of achene level and storage. *J Agric Food Chem.* 55(13): 5156–5166. <https://doi.org/10.1021/jf070467u>
- Amoriello T., Ciccoritti R. and Ferrante P. 2022. Prediction of strawberries' quality parameters using artificial neural networks. *Agronomy.* 12(4):963. <https://doi.org/10.3390/agronomy12040963>
- Athira K., Sooraj N.P., Jaishanker R., Saroj Kumar V., Sajeev C.R., Pillai M.S., Govind A. and Dadhwal V.K. 2019. Quantitative representation of floral colors. *Color Res Appl.* 44(3):426–432. <https://doi.org/10.1002/col.22353>
- Barbin D.F., Mastelini S.M., Barbon S., Campos G.F.C., Barbon A.P.A.C. and Shimokomaki M. 2016. Digital image analyses as an alternative tool for chicken quality assessment. *Biosyst Eng.* 144:85–93. <https://doi.org/10.1016/j.biosystemseng.2016.01.015>
- Bilgili M. and Sahin B. 2010. Comparative analysis of regression and artificial neural network models for wind speed prediction. *Meteorol Atmos Phys.*, 109:61–72. <https://doi.org/10.1007/s00703-010-0093-9>
- Buvé C., Kebede B.T., De Batselier C., Carrillo C., Pham H.T.T., Hendrickx M., Grauwet T. and Van Loey A. 2018. Kinetics of colour changes in pasteurised strawberry juice during storage. *J. Food Eng.* 216:42–51. <https://doi.org/10.1016/j.jfoodeng.2017.08.002>
- Caner C., Aday M.S. and Demir M. 2008. Extending the quality of fresh strawberries by equilibrium modified atmosphere packaging.

- Eur Food Res Technol. 227(6):1575–1583. <https://doi.org/10.1007/s00217-008-0881-3>
- Contreras-López E., Jaimez-Ordaz J., Ugarte-Bautista I., Ramírez-Godínez J., González-Olivares L.G., García-Curiel L. and Pérez-Flores J.G. 2022. Use of image analysis to determine the shelf-life of an apple compote with wine. *Food Sci Technol.* 42:e04122. <https://doi.org/10.1590/fst.04122>
- Da Silva Simão R., De Moraes J.O., Lopes J.B., Frabetti A.C.C., Carciofi B.A.M. and Laurindo J.B. 2022. Survival analysis to predict how color influences the shelf life of strawberry leather. *Foods.* 11(2):218. <https://doi.org/10.3390/foods11020218>
- Galoburda R., Boca S., Skrupskis I., Seglina D. 2014. Physical and chemical parameters of strawberry puree. 9th Baltic Conference on Food Science and Technology “Food for Consumer Well-Being,” FOODBALT Conference Proceedings; Jelgava, Latvia, May 8–9, 2014, pp. 172–177.
- Hernanz D., Recamales Á.F., Meléndez-Martínez A.J., González-Miret M.L. and Heredia F.J. 2008. Multivariate statistical analysis of the color–anthocyanin relationships in different soil less-grown strawberry genotypes. *J Agric Food Chem.* 56(8):2735–2741. <https://doi.org/10.1021/jf073389j>
- Hosseinifarahi M., Jamshidi E., Amiri S., Kamyab F. and Radi M. 2020. Quality, phenolic content, antioxidant activity, and the degradation kinetic of some quality parameters in strawberry fruit coated with salicylic acid and Aloe vera gel. *J Food Process Preserv.* 44(9):e14647. <https://doi.org/10.1111/jfpp.14647>
- Howard L.R., Brownmiller C. and Prior R.L. 2014. Improved color and anthocyanin retention in strawberry puree by oxygen exclusion. *J Berry Res.* 4(2):107–116. <https://doi.org/10.3233/JBR-140072>
- Kafkas E., Koşar M., Paydaş S., Kafkas S. and Başer K.H.C. 2007. Quality characteristics of strawberry genotypes at different maturation stages. *Food Chem.* 100(3):1229–1236. <https://doi.org/10.1016/j.foodchem.2005.12.005>
- Kato T., Mastelini S.M., Campos G.F.C., Da Costa Barbon A.P.A., Prudencio S.H., Shimokomaki M., Soares A.L. and Barbon S. 2019. White striping degree assessment using computer vision system and consumer acceptance test. *Asian-Australasian J Anim Sci.* 32(7):1015–1026. <https://doi.org/10.5713/ajas.18.0504>
- Kazi M.K., Eljack F. and Mahdi E. 2021. Data-driven modeling to predict the load vs. displacement curves of targeted composite materials for industry 4.0 and smart manufacturing. *Compos Struct.* 258:113207. <https://doi.org/10.1016/j.compstruct.2020.113207>
- Lev A., Braw Y., Elbaum T., Wagner M. and Rassovsky Y. 2022. Eye tracking during a continuous performance test: utility for assessing ADHD patients. *J Atten Disord.* 26(2):245–255. <https://doi.org/10.1177/1087054720972786>
- Liu C., Zheng H., Sheng K., Liu W. and Zheng L. 2018. Effects of melatonin treatment on the postharvest quality of strawberry fruit. *Postharvest Biol Technol.* 139:47–55. <https://doi.org/10.1016/j.postharvbio.2018.01.016>
- Lopes J.F., Ludwig L., Barbin D.F., Grossmann M.V.E. and Barbon S. 2019. Computer vision classification of barley flour based on spatial pyramid partition ensemble. *Sensors (Switzerland).* 19(13):2953. <https://doi.org/10.3390/s19132953>
- Lukawska-Matuszewska K. and Urbański J.A. 2014. Prediction of near-bottom water salinity in the Baltic Sea using ordinary least squares and geographically weighted regression models. *Estuar Coast Shelf Sci.* 149:255–263. <https://doi.org/10.1016/j.ecss.2014.09.003>
- Lv J., Zheng T., Song Z., Pervaiz T., Dong T., Zhang Y., Jia H. and Fang J. 2022. Strawberry proteome responses to controlled hot and cold stress partly mimic post-harvest storage temperature effects on fruit quality. *Front Nutr.* 8:812666. <https://doi.org/10.3389/fnut.2021.812666>
- Mancini M., Mazzoni L., Gagliardi F., Balducci F., Duca D., Toscano G., Mezzetti B. and Capocasa F. 2020. Application of the non-destructive NIR technique for the evaluation of strawberry fruits quality parameters. *Foods.* 9(4):441. <https://doi.org/10.3390/foods9040441>
- Marszałek K., Mitek M. and Skąpska S. 2015. Effect of continuous flow microwave and conventional heating on the bioactive compounds, colour, enzymes activity, microbial and sensory quality of strawberry purée. *Food Bioprocess Technol.* 8(9):1864–1876. <https://doi.org/10.1007/s11947-015-1543-7>
- Mollalo A., Vahedi B. and Rivera K.M. 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci Total Environ.* 728:138884. <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Obradović V., Ergović Ravančić M., Marčetić H. and Škrabal S. 2020. Properties of strawberries puree stored in the freezer. *Ital J Food Sci.* 32(4):945–955. <https://doi.org/10.14674/IJFS.1858>
- Oliveira M.M., Cerqueira B.V., Barbon S. and Barbin D.F. 2021. Classification of fermented cocoa beans (cut test) using computer vision. *J Food Compos Anal.* 97:103771. <https://doi.org/10.1016/j.jfca.2020.103771>
- Oszmiański J., Wojdyło A. and Kolniak J. 2009. Effect of l-ascorbic acid, sugar, pectin and freeze-thaw treatment on polyphenol content of frozen strawberries. *LWT.* 42(2):581–586. <https://doi.org/10.1016/j.lwt.2008.07.009>
- Palkovits S. 2020. A primer about machine learning in catalysis – a tutorial with code. *Chem Cat Chem.* 12(16):3995–4008. <https://doi.org/10.1002/cctc.202000234>
- Pilkington J.L., Preston C. and Gomes R.L. 2014. Comparison of response surface methodology (RSM) and artificial neural networks (ANN) towards efficient extraction of artemisinin from *Artemisia annua*. *Ind Crops Prod.* 58:15–24. <https://doi.org/10.1016/j.indcrop.2014.03.016>
- Prihanditya H.A. and Alamsyah. 2020, Oct. The implementation of Z-score normalization and boosting techniques to increase accuracy of C4.5 algorithm in diagnosing chronic kidney disease. *J Soft Comput Explor.* 1(1):63–69. <https://doi.org/10.52465/josce.v1i1.8>
- Rogiers B., Mallants D., Batelaan O., Gedeon M., Huysmans M. and Dassargues A. 2012. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Math Geosci.* 44(6):739–763. <https://doi.org/10.1007/s11004-012-9409-2>
- Santos Pereira L.F., Barbon S., Valous N.A. and Barbin D.F. 2018. Predicting the ripening of papaya fruit with digital

- imaging and random forests. *Comput Electron Agric.* 145:76–82. <https://doi.org/10.1016/j.compag.2017.12.029>
- Scalisi A., O'connell M.G., Islam M.S. and Goodwin I. 2022. A fruit colour development index (CDI) to support harvest time decisions in peach and nectarine orchards. *Horticulturae.* 8(5):459. <https://doi.org/10.3390/horticulturae8050459>
- Stan A., Stan A. and Popa M.E. 2016. Pretreatment behavior of frozen strawberries and strawberry purees for smoothie production. *Sci Bull Ser F Biotechnol.* 19(February):315–323.
- Teribia N., Buvé C., Bonerz D., Aschoff J., Hendrickx M. and Loey A. Van. 2021. Impact of processing and storage conditions on color stability of strawberry puree: the role of PPO reactions revisited. *J Food Eng.* 294:110402. <https://doi.org/10.1016/j.jfoodeng.2020.110402>
- Tosun E., Aydin K. and Bilgili M. 2016. Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. *Alex Eng J.* 55(4):3081–3089. <https://doi.org/10.1016/j.aej.2016.08.011>
- Udomkun P., Nagle M., Argyropoulos D., Wiredu A.N., Mahayothee B. and Müller J. 2017. Computer vision coupled with laser backscattering for non-destructive colour evaluation of papaya during drying. *J Food Meas Charact.* 11(4):2142–2150. <https://doi.org/10.1007/s11694-017-9598-y>
- Ulrici A., Foca G., Ielo M.C., Volpelli L.A. and Lo Fiego D. Pietro. 2012. Automated identification and visualization of food defects using RGB imaging: application to the detection of red skin defect of raw hams. *Innov Food Sci Emerg Technol.* 16:417–426. <https://doi.org/10.1016/j.ifset.2012.09.008>
- Valous N.A., Mendoza F., Sun D.W. and Allen P. 2009. Colour calibration of a laboratory computer vision system for quality evaluation of pre-sliced hams. *Meat Sci.* 81(1):132–141. <https://doi.org/10.1016/j.meatsci.2008.07.009>
- Wagle P., Xiao X., Gowda P., Basara J., Brunzell N., Steiner J. and Anup K.C. 2017. Analysis and estimation of tallgrass prairie evapotranspiration in the central United States. *Agric For Meteorol.* 232:35–47. <https://doi.org/10.1016/j.agrformet.2016.08.005>
- Wu D. and Sun D.W. 2013. Colour measurements by computer vision for food quality control – a review. *Trends Food Sci Technol.* 29(1):5–20. <https://doi.org/10.1016/j.tifs.2012.08.004>
- Xu S., Wang J., Shou W., Ngo T., Sadick A.M. and Wang X. 2021. Computer vision techniques in construction: a critical review. *Arch Comput Methods Eng.* 28(5):3383–3397. <https://doi.org/10.1007/s11831-020-09504-3>
- Zhang L., Wang L., Zeng X., Chen R., Yang S. and Pan S. 2019. Comparative transcriptome analysis reveals fruit discoloration mechanisms in postharvest strawberries in response to high ambient temperature. *Food Chem X.* 2100025. <https://doi.org/10.1016/j.fochx.2019.100025>
- Zheng H., Jiang L., Lou H., Hu Y., Kong X. and Lu H. 2011. Application of artificial neural network (ANN) and partial least-squares regression (PLSR) to predict the changes of anthocyanins, ascorbic acid, total phenols, flavonoids, and antioxidant activity during storage of red bayberry juice based on fractal analysis and red, green, and blue (RGB) intensity values. *J Agric Food Chem.* 59(2):592–600. <https://doi.org/10.1021/jf1032476>