Using Machine Learning Algorithms to Predict the Sweetness of Bananas at Different Drying Times

Sufyan A. Al-Mashhadany¹, Haider Ali Hasan¹, Mustafa A.J. Al-Sammarraie*¹

¹ Department of Agricultural Machinery and Equipment, University of Baghdad, College of Agricultural Engineering Sciences, Baghdad, Iraq
* Corresponding author's e-mail: mustafa.ahm@coagri.uobaghdad.edu.iq

ABSTRACT
The consumption of dried bananas has increased because they contain essential nutrients. In order to preserve bananas for a longer period, a drying process is carried out, which makes them a light snack that does not spoil quickly. On the other hand, machine learning algorithms can be used to predict the sweetness of dried bananas. The article aimed to study the effect of different drying times (6, 8, and 10 hours) using an air dryer on some physical and chemical characteristics of bananas, including CIE-L*a*b, water content, carbohydrates, and sweetness. Also predicting the sweetness of dried bananas based on the CIE-L*a*b ratios using machine learning algorithms RF, SVM, LDA, KNN, and CART. The results showed that increasing the drying time led to an increase in carbohydrates, sweetness, and CIE-L*a*b levels, while it led to a decrease in the moisture content in dried banana slices. Therefore, there is a direct relationship between CIE-L*a*b levels and sweetness. On the other hand, the RF and CART algorithms gave the highest prediction accuracy of 86% and 0.8 on the Kappa measure. While the other algorithms (SVM, LDA, KNN) gave a prediction accuracy of 80% and 0.7 on the Kappa measure. In terms of testing statistical significance, the null hypothesis (H0) was accepted because there is no relationship between the metric distributions of the algorithms used.

Keywords: drying time, machine learning, prediction, sweetness, quality.
by temperature and relative humidity. Research shows that the formation of brown color in dried fruits increases with increasing temperature and decreasing humidity (Güzel et al., 2022). Baini and Langrish (2009) dried bananas at 70 degrees Celsius and relative humidity from 15% to 40% and found that the change in redness of the sample dried at 15% relative humidity was 67% higher than that of the sample dried at 40% relative humidity. Drying time also plays an important role in influencing the color of bananas, as research has shown that controlling drying time is crucial to maintaining the desired color properties, as the interaction between drying time, temperature, and humidity affects color changes during drying (Pekke et al., 2013). The drying process affects the sweetness of the fruits by affecting the sugar concentration during the drying of the fruits, that is, the longer the drying period, the more water evaporates from the fruit, and thus the sugar concentration increases. The exact quantitative relationship depends on factors such as the type of fruit, the initial sugar content, and the drying method used (Leite et al., 2007). On the other hand, during the drying process, the carbohydrate content in dried fruits is concentrated, because when more water is removed, more sugars are concentrated (Correia et al., 2009). Therefore, drying the fruit helps preserve the carbohydrates and other nutrients present in the banana (Alagbe et al., 2020). When performing fruit sorting operations, it is not easy to distinguish between the quality levels of fruit available in the market. This manual process requires a lot of work and therefore increases costs and time. On the other hand, the manual sorting process affects the quantity of exports due to poor sorting and classification of fruit. The use of automated sorting systems based on machine vision can improve fruit quality, eliminating inconsistencies caused by manual evaluation. Researchers have already begun to use machine learning algorithms to perform a wide range of calculations and analyze data that lead to relevant discoveries that can help improve the fruit sorting and grading process. As a result, different aspects of the fruit are being studied using machine learning techniques. These aspects include color, texture, quality, taste, etc. Al-Sammarraie et al., (2022) predicted the sweetness of orange fruits by studying the relationship between RGB values and fruit sweetness using different machine-learning algorithms. Ma et al., (2022) also predicted the ripeness of bananas by evaluating the fruit’s color and sweetness using the RF, ANN, and SVM algorithms. Amoriello et al., (2022) characterized the physicochemical and nutritional characteristics of seven strawberry cultivars at different harvest times and verified the effectiveness of MLR and ANN algorithms to build models to predict these traits using color space coordinates. Kondo et al., (2000) used machine vision and neural network technology to evaluate the quality of Iyokan orange fruit to automate the orange fruit grading process. Mazen and Nashat (2019) presented a new effective algorithm to determine banana ripeness, as the proposed technique is based on HSV color, brown spot development, and banana fruit texture analysis. By analyzing data and using predictive models, it is possible to predict the sweetness of fruits based on their physical and chemical characteristics. The article aimed to study the effect of different drying times (6, 8, and 10 hours) using an air dryer on some physical and chemical characteristics of bananas, including CIE-L*a*b, moisture content, carbohydrates, and sweetness. Also predicting the sweetness of dried bananas based on the CIE-L*a*b ratios using machine learning algorithms RF, SVM, LDA, KNN, and CART.

MATERIALS AND METHODS

Sample preparation

The experiment was conducted on samples of banana slices collected from local markets (Iraq, Baghdad) of the Cavendish variety. Banana samples were chosen to be of equal size and degree of maturity. Table 1 shows the characteristics of the banana slices used in the experiment.

Measurement methodology

Machine learning algorithms were used in this research, more specifically the RF, SVM, LDA, KNN, and CART algorithms to predict the sweetness of bananas after the drying process. The research methodology consists of four steps: The first step is to conduct the process of drying banana slices at three different times (6, 8, and 10 hours). The second step is to take data on the sensory characteristics of banana slices after the drying process, including moisture content, sweetness, carbohydrates, and color. The third step is to determine the relationship between drying time
and measured sensory properties in general and the relationship between CIE-L*a*b values and sweetness percentage. The final step was to identify the most accurate prediction algorithms.

Drying process

In this experiment, an electric food dehydrator from Sooba was used, a cylindrical shape consisting of 5 trays, equipped with a temperature thermostat that is adjustable between 8 levels from 95 to 158 degrees Fahrenheit. The wattage was 250 watts, and the voltage was 230 volts. After cutting them, the banana slices were placed in the device’s trays at a temperature of 70 degrees Celsius. The drying process was carried out at three different times (6, 8, and 10 hours). The humidity of the drying device affects the speed of drying of fruits. The evaporation rate is affected by relative humidity to a greater extent than temperature (Chanpet et al., 2020). Therefore, the WBGT device was used to measure the humidity inside the dryer during the drying process. The humidity reached 60%. The specific humidity is 0.005 g water g⁻¹ dry air.

Measuring sensory properties

To measure the color contrast in banana slices before and after the drying process, use a fiber optic sensor. The working principle of the sensor is based on sending light at an angle of 10 to the surface of the banana slice, then measuring the amount of light reflected from the surface of the slice after the reflected light passes through 31 filters to measure the entire color spectrum. The measured spectral reflectance data are easily converted into three numbers representing CIE-L*a*b (Al-Sammarraie et al., 2023). A pocket molecular scanner SCiO (Consumer Physics Ltd., Illinois, USA) was also used to measure sugar (Brix), carbohydrates, and water content with an accuracy of 78% before the drying process (Jihad et al., 2024). After drying bananas, the SCiO device used to measure the percentage of sweetness, carbohydrates, and water content, which relies on the reflection of light, gives inaccurate measurements. This is because dried bananas change their optical properties, which affect how light interacts with their surface. Therefore, the above characteristics were measured using the laboratory method.

Moisture content was determined by evenly distributing 10 g dried banana samples onto pre-weighed aluminum dishes and drying them for at least 6 h at 105 °C using an electric oven. The moisture content was determined based on the initial and final sample weights using the following equation (Abdulkader et al., 2021; Kareem and Jasim, 2023):

$$ MC_{db} = \frac{M_i - M_f}{M_f} $$ (1)

where: $C_{db}$ – moisture content is dry basis; $M_i$ – the initial mass (g); $M_f$ – the dry sample mass (g).

The total color change was then determined using the following equation (Pekke et al., 2013):

$$ \Delta E = \sqrt{(L_i^* - L_f^*)^2 + (a_i^* - a_f^*)^2 + (b_i^* - b_f^*)^2} $$ (2)

where: $\Delta E$ – a total color change, $L_i^*$, $a_i^*$, $b_i^*$ – are initial color values, and $L_f^*$, $a_f^*$, $b_f^*$ – final color values.

The amount of carbohydrates in dried bananas was also estimated by calculating the difference between the original sample weight and the sum of the weights of the ether extract, crude protein, crude fiber, and ash, and based on the following equation (Cömert et al., 2015; Kareem and Shakir, 2016):

$$ \text{NFE\%} = \text{DM\%} - (\text{EE} + \text{CP} + \text{CF} + \text{Asb}) $$ (3)


To evaluate the sweetness of banana slices after the drying process, we relied on the sensory evaluation method. Samples of dried banana slices were presented to 50 faculty members and students of the Department of Agricultural Machinery and Equipment at the University of Baghdad.

Data processing

This experiment was conducted using the R programming program, which contains various
office functions that can be used to implement different algorithms. The software is installed on a Lenovo device with 16 GB of storage, and a 2.3 GHz Intel Core i7 processor. This dataset contains three levels of sweetness for bananas. The total data obtained was 150 for three different drying times. Data preprocessing was performed, by normalizing unwanted data such as missing data, random data, and noise. We split the dataset into smear sizes to understand the result. Recursive cross-validation is used as a common configuration criterion, where we use 10 folds with 3 iterations. Rely on the accuracy and kappa scale to evaluate the used algorithms. Accuracy here is the percentage of correct classification based on the total classified instances. Kappa is normalized to the baseline for random selection of the data set. For easy classification of banana slices, the inputs included the CIE-L*a*b values of banana slices after drying while the outputs included the sweetness level. To compare different network architectures, the training and test sets were randomly selected to be used for all tests in this study.

RESULTS AND DISCUSSION

The results indicate that the moisture content is affected by the drying time, as the amount of moisture removed from the sample increases with increasing drying time, as explained in Figure 1. The moisture percentage after 6 hours of drying was 34.39%, 20.82% after 8 hours of drying, and 16.94% after 10 hours of drying. Based on the above results, the percentage of moisture removed when the drying time was increased from 6 to 8 hours was 39.35% and 18.63% when the drying time was increased from 8 to 10 hours. This indicates that the time the samples remained inside the dryer led to exposing the sample to more heat, which led to the removal of the largest amount of moisture inside the sample (Sallam et al., 2013). Also, the drying rate is highest in the first drying hours and decreases with time because of the decrease in internal resistance to moisture at the beginning of drying (Chininye, 2009).

It is clear from Figure 2 that the lowest amount of carbohydrates was obtained at a drying time of 6 hours, amounting to 80.07 mg g⁻¹, while the amount of carbohydrates at the minimum drying time of 8 hours amounted to 87.12 mg g⁻¹ and 153.66 mg g⁻¹ at Drying time 10 hours. The carbohydrate concentration increased by 8.8% when the drying time increased from 6 to 8 hours, while the concentration increased by 76.37% when the drying time increased from 8 to 10 hours. The above results indicate a direct relationship between carbohydrate content and drying time, as drying time increases, the concentration of carbohydrates in samples dried at a constant temperature increases. Based on the sensory evaluation method, 50 samples of dried bananas were tasted at different drying times. 42 people agreed that banana samples dried at a drying time of 10 hours gave the highest sweetness, while 8 people confirmed that bananas dried at 8 hours did not differ

![Figure 1. Relationship of moisture content to drying time](image-url)
in sweetness from bananas dried at a drying time of 10 hours. Also, all people agreed that banana samples dried at a drying time of 6 hours are the least sweet. The validity of the sensory evaluation is that the concentration of carbohydrates increases with increasing drying time and that carbohydrates are directly proportional to the percentage of sugar, as the latter is related to the percentage of sweetness.

From Figure 3, the results show that there is a direct relationship between the CIE-L*a*b values with increasing drying time. As the processing time increases, the CIE-L*a*b values increase, making it a light yellowish color (Prieto-Santiago et al., 2015). The brown color of the samples was obtained at the shortest drying time. This is because samples dried at a drying time of 6 hours contain the highest percentage of moisture content and the water present in the sample is one of the reasons for the formation of the brown color (Baini and Langrish, 2009).

Regarding color change, the drying time of 10 hours gave the lowest color change value of 6.85. The drying time of 6 hours gave the highest color change value of 7.79, and the drying time of 8 hours gave a value of 7.22. This is consistent with the results obtained, as the drying time of 10 hours maintained the color close to the natural color of bananas before drying. Dried banana slices were classified based on machine learning algorithms and based on the relationship between sweetness and color. The amount of sweetness of dried banana samples increased with increasing drying time, and at the same time, the levels of CIE-L*a*b increased, meaning that there is a direct relationship between sweetness and CIE-L*a*b.

In terms of evaluating the accuracy of the algorithms used in the experiment using the R
programming program, Figure 4 taken from the program above, shows a summary of the algorithm data, which includes the minimum, median, and maximum values, and some percentages for each measure (Alagbe et al., 2020). The CART and RF algorithms gave the highest prediction accuracy value of 86%. While the other algorithms (SVM, LDA, KNN) gave a prediction accuracy of 80%. The CART and RF algorithms also excelled in terms of the Kappa scale, giving an accuracy of 0.8, while the other algorithms gave an accuracy of 0.7. The accuracy of the CART and RF algorithms based on the Kappa measure showed very good agreement, while the other algorithms gave good agreement. Kappa values less than 0.20 are considered poor agreement, while values ranging between 0.8–1.0 are considered very good agreement (Gupta et al., 2022).

The distribution of data in Figure 5 can be checked based on skewness. Skewness is the degree to which data is distorted from the central value r. This method is useful for looking at the spread of the estimated accuracy of different algorithms and how they are related. Average values are represented in the figure by dots overlapping the boxes. It is clear from Figure 5 that the SVM and LDR algorithms have a larger normal distribution compared to other algorithms.

For statistical significance tests, the significance of the differences between the metric distributions of the machine learning algorithms used in the research was calculated. Figure 6 shows the pairwise statistical significance scores, the bottom
diagonal of the table shows the p-values, and the top diagonal of the table shows the variation between the distributions. The P value is used to indicate the probability of uncorrelated systems (p < 0.05). The null hypothesis was addressed: H0: There is no relationship between CART, LDA, SVM, KNN, and RFK, and H1: There is a relationship between CART, LDA, SVM, KNN, and RFK. Based on the data in Figure 6, we reject hypothesis H1 and accept the null hypothesis H0.

CONCLUSIONS

Increasing the drying time from 6 to 8 hours led to an increase in carbohydrates by 8.8% and an increase of 76.37% when increasing the drying time from 8 to 10 hours. The drying time from 6 to 8 hours led to a reduction in moisture content by 39.35% and by 18.63% when the time was increased from 8 to 10 hours. Also, increasing drying time led to increased CIE-L*a*b levels. The amount of color change at a drying time of 6 was 6.85 and at 8 hours was 7.22, while the amount of color change reached the lowest level at a drying time of 10 hours of 7.79. There is a direct relationship between drying time, sweetness, and CIE-L*a*b levels. To predict the sweetness of dried bananas, the RF and CART algorithms gave the highest prediction accuracy of 86% and 0.8 on the Kappa measure. While the other algorithms (SVM, LDA, KNN) gave a prediction accuracy of 80% and 0.7 on the Kappa measure. The null hypothesis (H0) was accepted because there is no relationship between the metric distributions of the algorithms used. Predicting the sweetness of dried bananas using machine learning algorithms gives more accurate measurements, improves the drying process, and also contributes to the development of the classification system. These benefits can improve product quality, enhance efficiency, and increase dried banana sales.

REFERENCES