Master's Thesis of Engineering

# Quality Estimation of Onions during Storage Periods using Machine Learning Techniques

머신 러닝 기법을 이용한 양파 저장기간 중 품질 평가에

관한 연구

February 2023

College of Agriculture and Life Science

Seoul National University

Biosystems Engineering

NANDITA IRSAULUL NURHISNA

# Quality Estimation of Onions during Storage Periods Using Machine Learning Techniques

**Advisor: Ghiseok Kim**

**Submitting a master's thesis of**

**Biosystems Engineering**

**February 2023**

**College of Agriculture and Life Science**
**Seoul National University**
**Biosystems Engineering**

**NANDITA IRSAULUL NURHISNA**

**Confirming the master's thesis written by**
**NANDITA IRSAULUL NURHISNA**
**February 2023**

Chair ______Joong Yong Rhee______(Seal)

Vice Chair ______Ghiseok Kim______(Seal)

Examiner ______Eun Jin Lee______(Seal)

# Abstract

Onions are a major vegetable in Korea. Long-term storage is therefore required to accommodate the demand throughout the year. Hence, storage needs to be considered carefully to extend the shelf life of onions. Temperature and relative humidity in storage plays a significant role in changing the quality of the onions, so temperature and humidity control during storage should be done to maintain the quality of the onions. Mechanical properties, weight loss, and respiration rate were chosen as the quality attributes of onions observed for quality changes during storage. In addition, developing a prediction model for changes in the quality of onions using machine learning needs to be carried out considering previous research, which is limited and only uses chemical kinetic models to predict changes in the quality of onions in storage.

In this study, we stored onion at 0-1℃, collected the environmental data, and did weekly destructive measurements for 10 weeks of storage periods from March to June 2022. We measured Bio-yield stress using a compression test, respiration rate, and weight loss based on a weight scale sensor installed inside the chamber. Based on the data collection, we constructed three machine learning models to make a quality estimation model for onion bio-yield strength and weight loss using environment data – time, temperature, and relative humidity. We used two datasets for bio-yield stress data with 100 data of 10 weeks measurement and 127 data of the augmentation dataset using polynomial interpolation degree 2. The machine learning technique used in this study were multiple linear regression (MLR), partial square-least regression (PLSR), and support vector regression

(SVR). The data were divided into train and test datasets in a ratio of 80:20 with 10-fold cross-validation on the training dataset. Then the regression models were evaluated by coefficient determination ($R^2$), root mean square regression error (RMSE), and mean absolute percentage error (MAPE).

From our study, the bio-yield stress decreased along with time, but the weight loss showed an increasing trend, for the respiration rate shows a relatively same trend since onion is a non-climacteric type. Furthermore, for the quality estimation model, we reported that the SVR and MLR models could be used to predict the quality attributes of onions during storage with $R^2$ values of >0.8 for bio-yield stress and $R^2$ >0.99 for weight loss parameters.

**Keywords:** Net packaged onion, wireless sensor network, quality estimation, machine learning technique, mechanical properties.

**Student Number: 2020-20878**

## Table of Contents

# List of Tables

# List of Figures

# Chapter 1. Introduction

## 1.1 Background

Onion (*Allium cepa L.*) is a vital crop worldwide because it has a variety of healthy ingredients like fiber, vitamins, organic acids, phenolic compounds, and other antioxidants (Sasongko *et al.*, 2020). (BAEK and KIM, 2020) stated that the cultivation of onion has increased by 4% over the last several years. Onions are generally planted in the fall and harvested from April to June of the following year, and then long-term storage is required to satisfy year-round supply. During long-term (almost one year) storage, onions are stored in a cold room which is controlled near $0^{o}C$, along with well-ventilated storage for longer shelf life. In Korea, onions are distributed mainly in a net packaging state after harvest. Onions in net packaging help the onions breathe more than in other closed-package systems. Also, the net packaging is made of rigid material, making it easy for workers to carry it. However, such a net packaging-based transport system may cause damage to onions. In addition, when rotten onions are found in the process, separating and removing them is difficult. However, harvesting happens primarily during the rainy season, and the high moisture content of onions can cause quality deterioration during long-term storage (Cho, Bae and Lee, 2010; Sang *et al.*, 2014). Moreover, physiological changes in onion bulbs during long-term storage can be affected by increased respiration, ethylene production, and other chemicals compound during the storage.

It was widely known that well-conditioned storage plays the primary role in delaying the deterioration of onions. Maintaining the optimum range of environmental conditions could prolong the onion's shelf life. Several studies have estimated onion quality and determined the optimal temperature during storage and distribution processes. To preserve the onion's quality, onions were kept at a low temperature of 0ºC and humidity around 60-75% (Jang and Lee, 2009; Porras-Amores, Mazarrón and Cañas, 2014; Isma'ila M, E. Karu, D. A Zhígila, 2017). Like this, temperature and humidity are important environmental factors in storage that would affect the freshness of onions. The quality of fruits and vegetables comprises multiple characteristics: sensory, nutritional, and mechanical properties (Abbott, 1999). Among those qualities of crops, mechanical property means how a crop behaves in response to an applied force. Several studies on the mechanical properties of potatoes, cucumbers, and apples during transportation and storage have been performed by (Masoudi, Tabatabaeefar and Borghaee, 2007; Eboibi and Uguru, 2017; Soliman and El-Sayed, 2017), it was evaluated that the external and internal forces can significantly influence mechanical damage in agricultural products. External forces are subjected to static and dynamic loads, resulting in injury, whereas internal forces can be caused by physical, chemical, and biological changes (Mohsenin, 2020). The onion's quality parameters, such as bio-yield stress (Ferreira *et al.*, 2015; Sharma *et al.*, 2015) and weight loss (Falayi, Yusuf and State, 2014; Emana *et al.*, 2017) were also studied to evaluate the quality change of onion for different storage conditions.

Recently, the field of quality control in agricultural products has employed advanced sensor network techniques such as innovative RFID and intelligent indicator packaging (Badia-Melis, Mishra and Ruiz-García, 2015; Chen, 2017; Xiao *et al.*, 2017; Shao *et al.*, 2021), and Wireless Sensor Network (WSN) is also used to measure the quality of the agricultural products in distribution or storage process. Some studies have suggested a high relationship between the quality of crops and the environmental data (temperature, humidity, and $CO_2$) of farmhouses that grow them (Mallik *et al.*, 2018; Sarmah and Aruna, 2020). Generally, the industrial supply chain monitors and controls the storage and distribution stages using a wireless sensor network in order to control the quality of distributed products (Chen, 2017; Accorsi et al., 2022). However, few studies evaluate onion quality based on real-time environmental data.

Since decades ago, many statistical or artificial learning-based methods have been used to evaluate or predict crop quality. Among these techniques, modeling techniques based on artificial neural networks have recently been widely applied (Chen, Ramaswamy and Alli, 2001; Ramzi *et al.*, 2015; Huang *et al.*, 2020; Correa-mosquera, Quicaz and Zuluaga-domínguez, 2022). The machine learning algorithm is a type of method that use historical data as input to predict new output values and become more accurate at predicting outcomes. In addition, supervised machine learning is implemented in these algorithms. Among the studies on onion quality prediction based on artificial neural networks, many used a kinetic model to predict the quality change in onions (Kaymak-Ertekin and Gedik, 2005; Devahastin and Niamnuy, 2010; Escobedo-Avellaneda *et al.*, 2012; Mitra, Shrivastava and Rao, 2015). However, these models have structural limitations in reflecting dynamic data, such as environmental information. Hence, in this study, we used a machine learning-based approach to develop a quality estimation of onions during storage using environmental information. Mainly we employed a Support Vector Regression method to estimate the quality of onions during storage since the method was known as one of the effective solutions for the problem of small sample size and non-linear attributes, which are precisely the quality attributes of onions during long-term storage.

## 1.2 Objectives

Mechanical properties deterioration causes nutrition loss and change in appearance. It is lowering the quality that meets the satisfactory consumer requirement. Based on the previous study storing the onion at inappropriate storage conditions would fasten the decay of onions. Therefore. This study intends to develop a quality estimation of onions during storage.

In this study, we measured the environmental temperature and humidity of the storage condition and observed the quality change (bio-yield stress, weight loss, and respiration rate) of the onions. Using these measurements, we developed and compared some quality estimation models using machine learning techniques such as multiple Linear regression, partial least square regression, and Support vector regression methods.

The main goal of this research is to establish and evaluate the precision of the onion's quality estimation that has been constructed using information regarding the conditions of storage environments. We created a simulated storage room that is similar to a general onion storage condition. The detailed goals to achieve the objectives are as follows:

1. Collect the environmental data of onion during storage periods
2. Analyze the onion's quality changes, such as bio-yield stress, weight loss, and respiration rate, that occur during cold storage.
3. Construct, analyze, and evaluate all three different types of machine learning models.

## 1.3 Literature review

The literature review will mainly focus on the research related to monitoring the quality of onions using a wireless sensor network in storing onions and quality estimations. In the research history about onion quality, we explain the quality of onions and what affected the onion quality after harvesting and during storage, the optimum storage environment for the onion to keep its shelf life. In wireless sensor network research history, we explain the current research using wireless sensor networks, the purpose of this method, and how it applies to our research. The last is about quality estimations. We explain possible models to use as quality estimations. Three regression models are explained: multiple linear regression, partial square least regression, and support vector regression. This part explains how these methods could be implemented as quality estimation in our research.

### 1.3.1 Onion quality

Quality onion after harvesting depends on their environmental situation. After harvesting the bulb, the onion should be cured first to reduce the moisture of the bulb and be kept at a low temperature to keep it for a longer time. Temperature and humidity play a significant role in decaying it. Several studies have been carried out to see a decrease in the quality of onions and find the optimal temperature for storage and distribution to maximize the storage of onions. (Jang and Lee, 2009) did research on the postharvest technology of onion, they compared several storage environments such as refrigerator condition, room temperature storage, and house storage. Keeping the onion at a higher temperature could avoid sprouting but will encourage decay. When storing the onion at a low temperature, both problems could be inhibited. (WARD, 1976) also stated that respiration increases with temperature. Water loss is a product of the respiration process. The higher the respiration, the more weight loss will occur. Even more important is the relative humidity of the storage. Onion differs from other horticultural commodities; onion requires a relative humidity of only 60-75% (Snowden, 1992). Moreover, onion has a dormant time; it is about 30-60 days after harvesting; no quality change will occur during that time due to sprouting growth.

The quality of onions in the distribution and storage process takes a significant loss if it did not store in an appropriate storage environment; several problems might occur during the storage and transportation of the bulb, such as high respiration, mold, and decay. Storage conditions play an essential role in the physiology of onions, which ultimately affects the physicochemical and phytochemical properties of onions and will be shown as loss of firmness and weight loss. The number of weeks in storage has indicated a decline in the weight loss of the bulb's shells, which also decreases the water content of the bulbs in storage. The quality of stored onions changes due to the high catabolism of substrates, primarily carbohydrates and other phytochemicals. Another indication of decayed that affected weight loss is sprouting (Isma'ila M, E. Karu, D. A Zhígila, 2017). Sprouting in bulbs increases in the number of storage weeks; the longer the sprouting, the bigger the weight declined. Sprouting also affected the weathering of the onion bulb and led to water loss. On the other hand, onions will have stacked each other during distribution and storage, affecting the onion's mechanical properties. Mechanical failure is classified as shear or clearage. (Brusewitz, McCollum and Zhang, 1991) Stated that excessive loads may also be the primary cause of bruising in fruits and vegetables.

The above studies show that the onion quality could be changed in storage due to environmental factors such as temperature and humidity. Several studies above also mentioned keeping the onion at a low temperature of about 0°C and in 60-75% humidity range to maintain the quality of the onion. Hence, in this research, we were trying to maintain the storing condition in the optimum condition and observed the quality change of onions, such as bio-yield stress, respiration rate, and weight loss.

### 1.3.2 Wireless sensor network

To reduce the quality loss of onions, it is best to have stored them at the optimum storage temperature and humidity of onion. The optimum storage environment for onions is near 0°C temperature and about 60-75% relative humidity. Keeping agricultural products fresh while stored is a vital concern for modern industry. Agricultural products could be considered fresh by controlling and monitoring the temperature and humidity. Temperature and humidity are critical for the storage of agricultural products since they significantly affect respiration and transpiration. Temperature and humidity regulation contribute to reducing degradation and prolonging the storage period. In the modern technology era, it is suggested to use a wireless sensor network to regulate and preserve the freshness of agricultural products in real time. Wireless Sensor Network detects and records the temperature and relative humidity from the sensor nodes. Collect and send the data to end-users via a wireless network (Correa et al., 2014).

Using a wireless sensor network provides easy transmission. It has many advantages over a traditional wire, better flexibility, and fast deployment features. Not only agricultural area but using wireless sensor networks also implemented in other areas such as the food cold chain (Aung and Chang, 2014b, 2014a), industry (Xu, Shen and Wang, 2014), healthcare (Hartley et al., 2018), and many others area. In agriculture itself, research in wireless sensor networks is developing and multipurpose. (Xiao et al., 2017) researched Wireless sensor networks to improve traceability and transparency in a cold chain of table grapes. WSN could determine the critical quality parameter of table grapes through real-time monitoring of the temperature fluctuation, resulting in the quality parameters of table grapes being affected by the temperature.

Wireless network sensors are widely applied in various agricultural applications. Using this method could also help maintain the freshness of onion during storage. Onions need to be kept in low storage and adequate relative humidity. A fluctuation in temperature and humidity in storage can cause the onion to decay and lose quality. Overall, in this study, we used a wireless sensor network to monitor the storage condition and assumed that storage condition and time would affect the quality of onions. We use the storage condition data as input data to predict the quality change in onions during storage.

### 1.3.3 Prediction model

Onion is rich in phytochemical compounds, which are claimed to have many health benefits. These compounds are sensitive to heat, oxygen, and light and will quickly degrade. In addition, nutritional compounds and color changes can directly affect consumer acceptance. The degradation of this critical value leads to the developing of a prediction model. A few prediction models have been proposed and tested. (Kaymak-Ertekin and Gedik, 2005) Reviewed the empirical method to predict a chemical change in the onion during the drying process.

Predictive analysis is the subfield of advanced analytics used to forecast future occurrences. Predictive analytics uses many techniques, from data mining, statistics, modeling, machine learning, and artificial intelligence, to analyze current data to make predictions. A predictive model works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes. (Melhem et al., 2016) Done a study to get a suitable method to predict the quality of wafers in the semiconductor manufacturing process based on production equipment data. As the semiconductor manufacturing process, it consists of a lot of correlated data and very few quality measurements. Regression models help build such a predictor using the production equipment data and quality measurements. Regression methods dealing with multicollinear high-dimensional input data are required.

A study on the prediction of quality changes in garlic cv was done by (Vázquez-Barrios et al., 2006). Principal components did correlation analysis, and a linear regression analysis was used to determine the degree of association of the parameters measured. The results indicate that the internal sprouting index and weight loss were the factors most affected by the storage conditions, and these factors could be used to predict shelf life.

A partial least squares regression (PLSR) is one of the popular methods in several studies. Using PLSR as a prediction analysis has been used in several studies. PLSR work to find the latent variable after decomposing X and Y. (Rozov, 2020) used PLSR to construct a predictive model in bioprocessing and found that it could predict the error and the robustness of Raman spectra. (Lee et al., 2018) found the difference between the Korean and Chinese soybean using PLSR, which was identified using transform infrared spectrometry (FT-IR). (Lim et al., 2014) demonstrated that NIR spectroscopy and a PLSR model could be helpful techniques for predicting rapidly and non-destructively the moisture content in red pepper powder.

The machine learning algorithm is a method that uses historical data as input to predict new output values and become more accurate at predicting outcomes. Supervised machine learning is implemented to predict continuous values. The SVR aims to achieve the optimal line inside a given threshold. Between the hyperplane and the borderline is the threshold value. SVR has an appropriate time complexity greater than quadratic with the number of samples, making it difficult to scale to datasets with more than a few thousand samples. (Sanaeifar, Bakhshipour and de La Guardia, 2016) Stated that SVR could effectively solve the small sample size. The concept of SVR is to map the input variables non-linearly into a high-dimensional feature set where they are linearly correlated with the output variable (Vapnik, 2000).

In conclusion, we construct the quality estimation of onions during storage periods using MLR for the simple linear regression, SVR for nonlinear regression, and PLSR to describe the input component that is simultaneously relevant to the output values using the PLSR method.

# Chapter 2. Materials and Methods

## 2.1 Sample and storage equipment

200 kg onions of the ‘Marusino 310’ variety were used as test material in this study. 'Marusino 310' is an extremely early-growing onion variety, which is sown in early September and harvested in early March at Jeollanam-do in a warm climate. The onion samples were transferred to a cold room at Seoul National University after harvesting without a curing process and stored in a range of 0-1°C as the optimum condition for storing onion (Jang and Lee, 2009). It is stored from March to June 2022 for three months period. The room storage used to store the onion is sized W1800×L1600×2000 mm (Hanbaek scientific corp. South Korea). Several sensors are set up inside the chamber to monitor environmental conditions in the storage room. Combined temperature, relative humidity sensors, and $CO_2$ (SH-VT-260-010, Sohatech, Republic of Korea) were put above the onions. Body temperature was inserted into two random bulbs about 30 mm. The air surface temperature probe was located between the bulb inside the boxes, and a weight scale (ES-30ki, A&D Korea Ltd, South Korea) was installed below 19 kg onions. A data acquisition device (Raspberry Pi 4B, Raspberry Pi Foundation, United Kingdom) was used to collect all sensor data every 30 minutes. Figure 1 shows how the sensors are installed inside the chamber, the specification of sensors used in this study is explained in Table 1, and a schematic of the monitoring systems is shown in Figure 2.

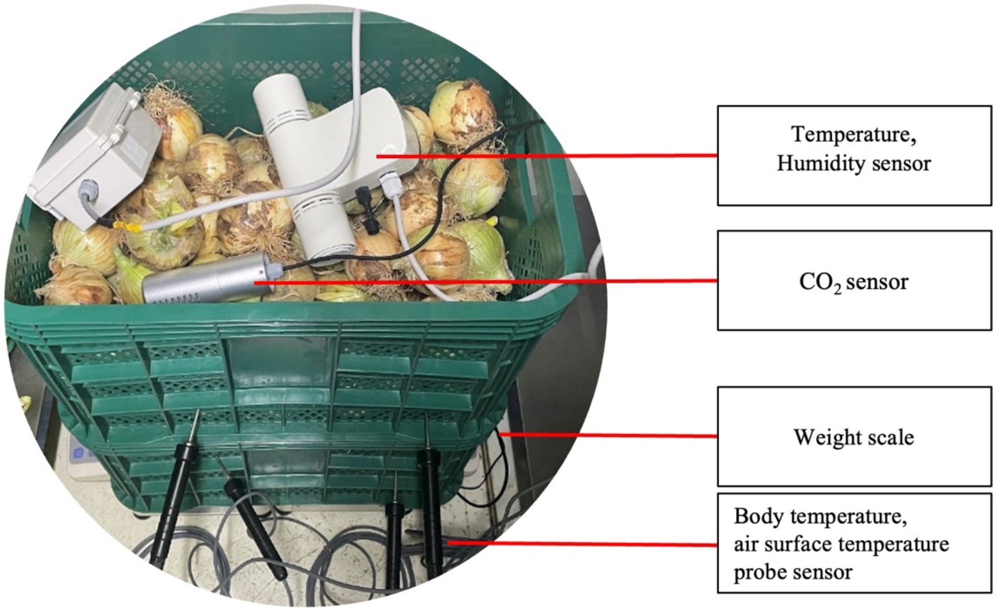



**Figure 1.** Pictures of setup sensors inside the room chamber.

**Table 1.** Specification of sensors installed in the chamber.

| **Sensor** | **Company** | **Type** | **Range** |
|---|---|---|---|
| Temperature | Soha tech. | SH-VT260 Rev D | -10.0~50.0°C |
| Relative humidity | Soha tech. | SH-VT260 Rev D | 0~99.9% |
| $CO_2$ | Soha tech. | SH-VT260 Rev D | 0~10000 ppm |
| Body temperature, air surface temperature | SYStronics | SCTS-06 | -45~80°C |
| Weight scale | A&D Korea Ltd | ES-30ki | Max 30 kg |

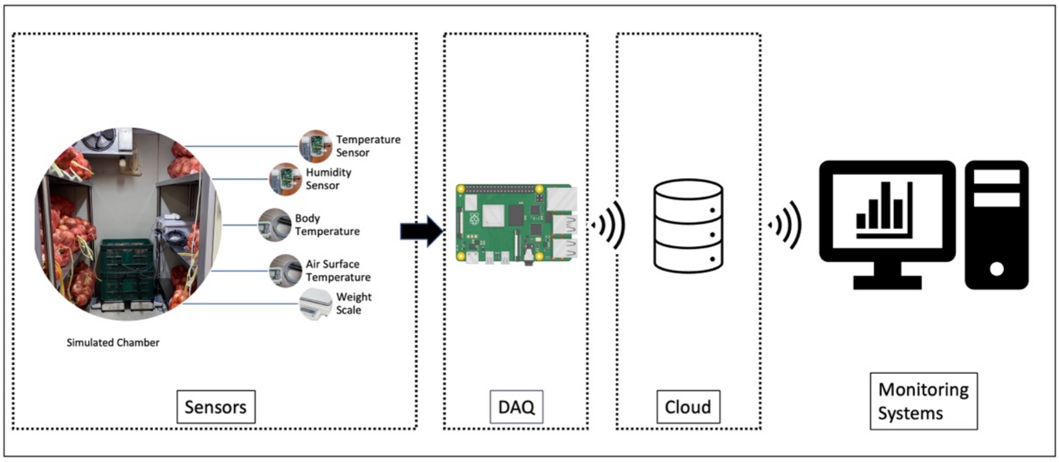



**Figure 2.** The schematic of monitoring systems used in this study.

## 2.2 Biophysical data measurement

Several destructive experiments were conducted weekly for a ten-weeks period to observe the deterioration of onion quality along with the storage time. Twenty-five onions were randomly sampled in each experiment; Twenty were used for mechanical property analysis, and the other five for respiration rate analysis.

### 2.2.1 Mechanical Properties

Mechanical property analysis was performed through a compression test using a 5 kN capacity universal testing machine (Autograph AGS-X series, Shimadzu Corp, Japan). The specification of the equipment is explained in Table 2. Before the compression, all onions were peeled and cut into half parallel to the face penetrating the roots and stems. The sample used and the compression method are shown in Figure 3 and Figure 4, respectively. We also observed the density of the onion sample using the water substitution method. It is expressed in Equation. (1). Then, the onion was placed in the machine above a flat plate, ensuring that the probe's center was aligned with the sample. Based on the standard for compression test of food material of convex shape, an 8 mm diameter probe and a speed rate of 25 mm/min was used (ASAE, 2008). Using the compression results, the bio-yield point was obtained as shown in Figure 5, and bio-yield stress was calculated based on Equation (2).

$$Density\ (kg/m^3) = \frac{W_d}{\gamma_w} \tag{1}$$

where $W_d$ is weight of sample (kg) and $\gamma_w$ is volum of sample inside the water

$$\sigma = \frac{F}{A} \tag{2}$$

where $\sigma$ is Stress ($N/mm^2$), F is Force (N), and A is Area ($mm^2$).

**Table 2.** Specification of universal testing machine.

| **Item** | **Specification** |
|---|---|
| Load cell capacity | Max. 5 kN |
| Crosshead speed | 0.001 mm/min to 1000 mm/min |
| Dimensions | W653 × D520 × H1603 mm |
| Power Supply Capacity | 1.2 kW |

**Figure 3.** Picture of peeled onion samples was used in the experiment.

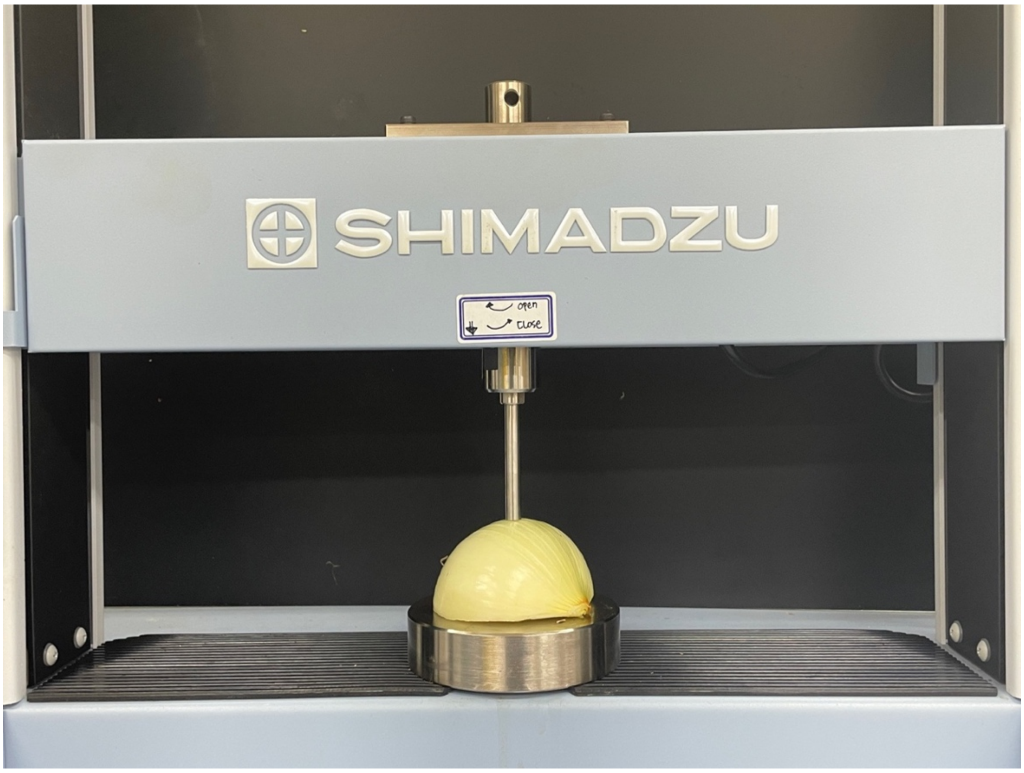



**Figure 4.** Picture of the compression test conducted using universal testing machine.

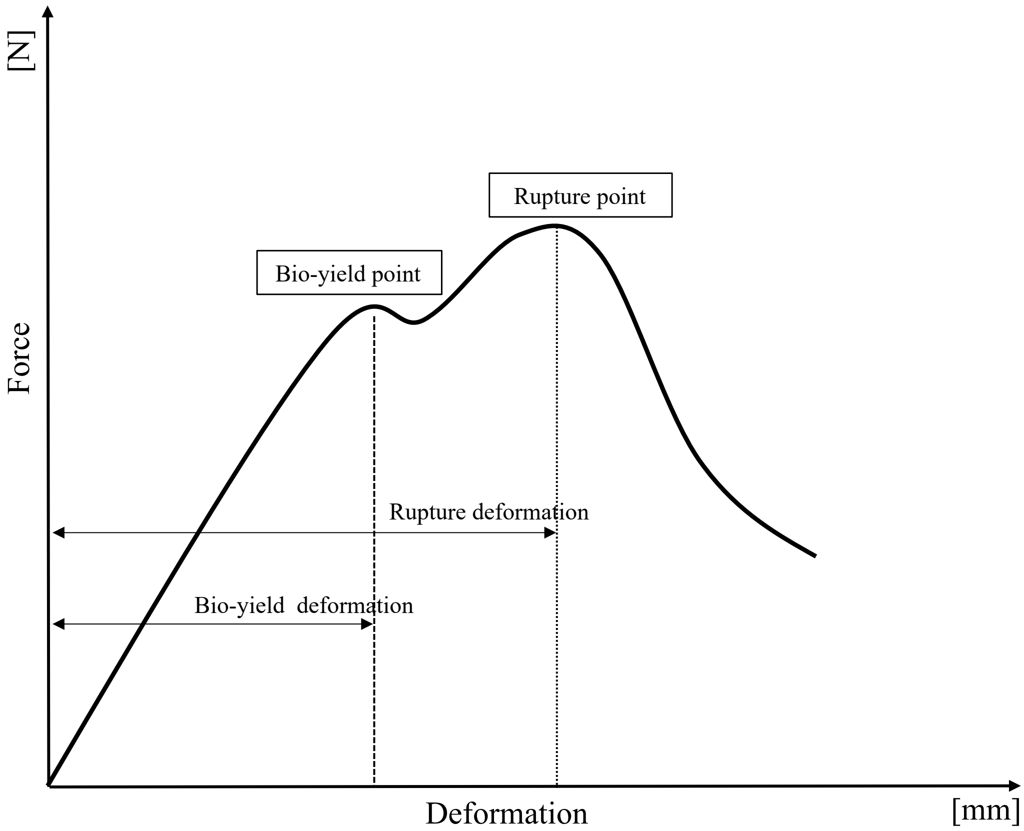



**Figure 5.** General force-deformation graph of a compression test.

### 2.2.2 Respiration rate

The respiration rate was calculated by measuring the amount of $CO_2$ emitted by onion samples at room temperature for 4 hours. To accurately detect the amount of $CO_2$, each sample was stored in a sealed 1L jar, and 10 ml gas sample was taken with a needle and injected into the gas Chromatography (6500GC, YL Instrument, Korea) (Figure 6 and Figure 7) the specification of the equipment is explained in Table 3. This instrument uses argon as a carrier gas at a flow rate of 18 ml/min and 50°C as a column temperature. The respiration of onion is expressed in Equation (3)

$$\mathrm{RCO_2} = \frac{V_f \times y\mathrm{CO_2}}{\mathrm{t} \times \mathrm{M}} \tag{3}$$

where, $\mathrm{RCO_2}$ (ml $CO_2$/kg/h) is the respiration rate, Vf (ml) is the free mass of the jar, $\mathrm{yCO_2}$ (decimal) is the volumetric concentration of $\mathrm{CO_2}$, t is the time that the sample stored in the room temperature, and M is the mass of the product (kg).

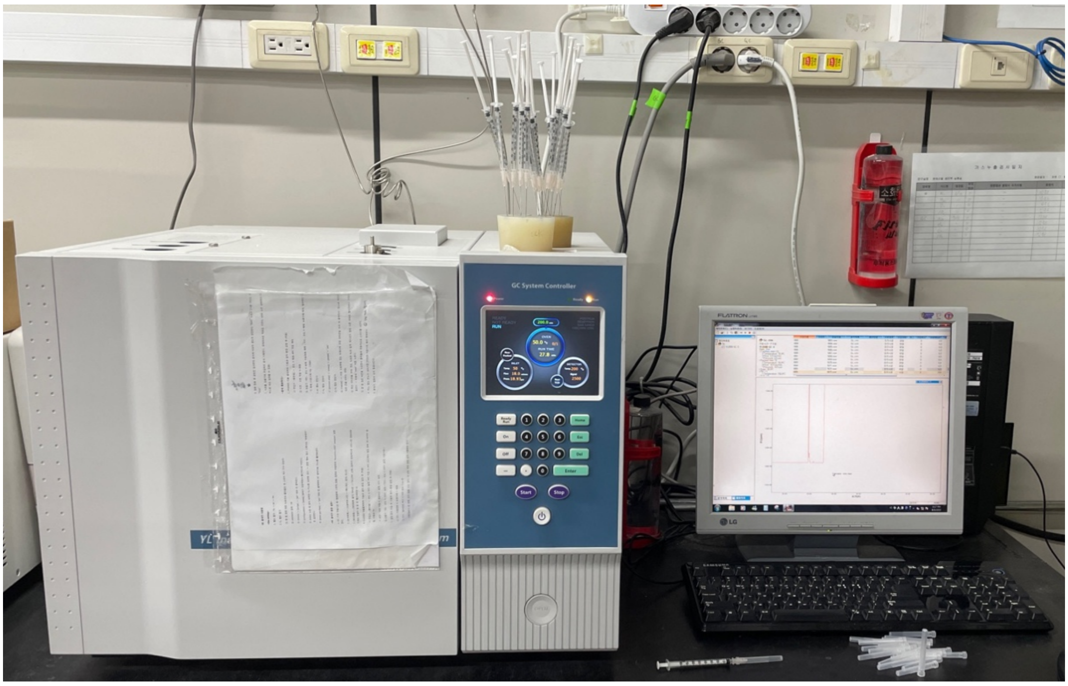


**Figure 6.** The picture of gas chromatography.

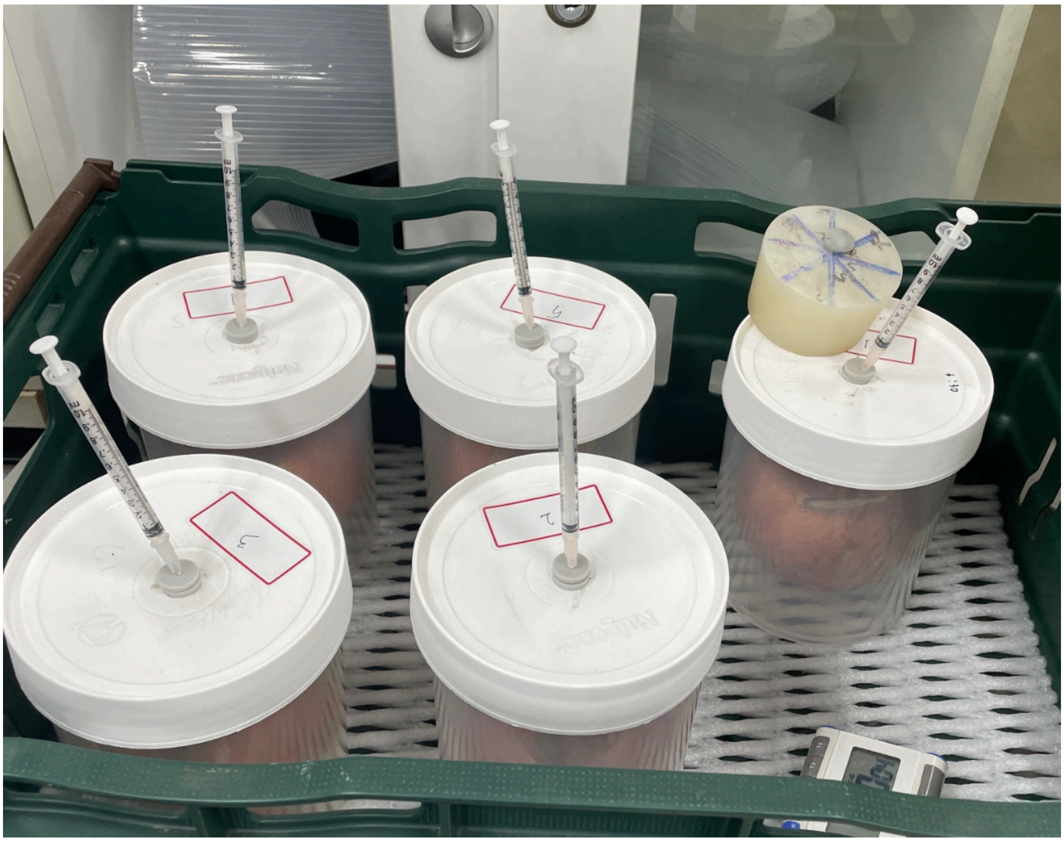


**Figure 7.** Picture of the $CO_2$ sampling technique.

**Table 3.** Specification of gas chromatography machine.

| **Item** | **Specification** |
|---|---|
| Carrier gas | Ar, 15 mL/min |
| Injector | 200℃ |
| Detector | TCD (200℃, FID (250℃), Methanizer 350℃ |
| Injection volume | 0.5 mL |
| Signal change | 4.0 min (TCD-FID) |

### 2.2.3 Weight loss

In order to quantitatively represent the weight loss, the weight loss rate and weekly weight change were calculated based on the value obtained from the scale installed in the cold room. The equations used in the calculation are shown below in Equations (4) and (5).

$$\text{Weight Loss Rate (g/kg)} = \frac{(\mathrm{W_0} - \mathrm{W_i})}{\mathrm{W_0}} \tag{4}$$

$$\text{Weight change (g/kg)} = \frac{(\mathrm{W_{(i+1)}} - \mathrm{W_i})}{\mathrm{W_0}} \tag{5}$$

where, $W_0$ is the initial weight and $W_i$ is the weight in i week.

It is known that most weight changes during storage occur from water loss, which can be inferred through the transpiration rate. The theoretical transpiration rate was calculated as shown in Equation (6) to compare with the measured weight change.

$$\mathrm{m} = \mathrm{k_t}(P_{SS} - P_{\infty}) \tag{6}$$

where, m represents a transpiration rate (g/kg s), $k_t$ is the transpiration coefficient (g/kg s Pa), $P_{ss}$ is water vapor pressure at the evaporating surface of the product (Pa), $P_{\infty}$ is ambient water vapor pressure (Pa). As a theoretical reference,

$850 \times 10^{-9}$g/kg s Pa was used as the transpiration coefficient value for onions (Sastry, 1985; Bovi *et al.*, 2016).

## 2.3 Quality estimation model

Based on the experiment, three types of quality estimation models based on machine learning techniques were developed to predict the bio-yield stress and weight loss of onions using the storage environment data - time, temperature, and relative humidity. The estimation models used in this study were multiple linear regression (MLR), partial least squares regression (PLSR), and support vector regression (SVR). For the development of the quality estimation models, a total of 100 experiment data were divided into train and test datasets in a ratio of 80:20, with 10-fold cross-validation on the training dataset (Figure 8). The regression models were evaluated by the coefficient of determination ($R^2$), root mean square error (RMSE), and mean absolute percentage error (MAPE), as described in Equations (7), (8), and (9), respectively. The regression analysis was performed in Python (Version 3.9.9).

$$R^2 = 1 - \frac{\sum(y_{act} - y_{pred})^2}{\sum(y_{act} - \bar{y}_{act})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum(y_{act} - y_{pred})^2}{n}} \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_{act} - y_{pred})}{y_{act}}\right| \times 100 \tag{9}$$

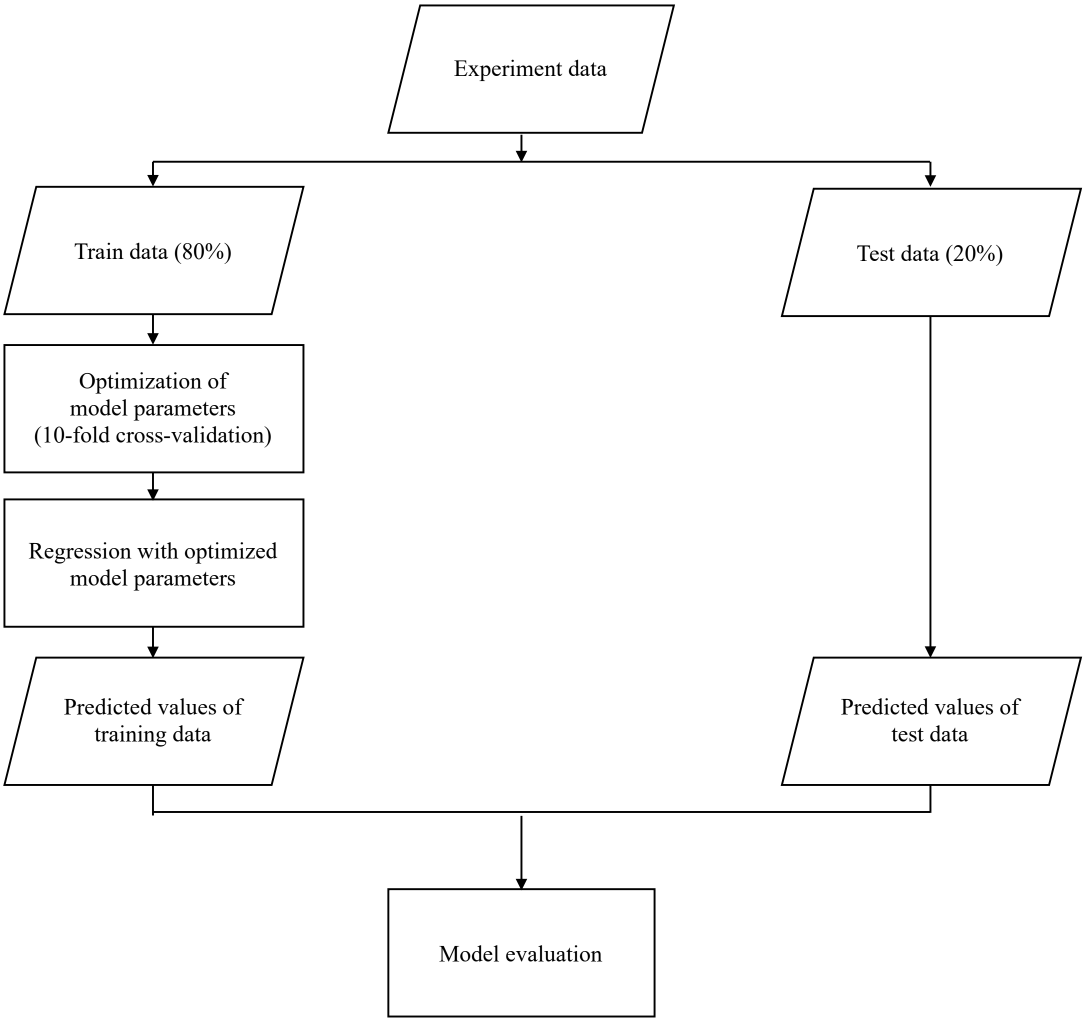



**Figure 8.** Flowchart of model training and evaluation.

### 2.3.1 Data Augmentation

Aside from that, we also perform data augmentation. This approach was taken to deal with the missing values of bio-yield stress data due to differences in the amount of data with environmental data: time, temperature, and relative humidity. To begin, we expanded the bio-yield stress data from 10 data of bio-yield stress each week average data to 127 data to match the length of sensor data using the polynomial interpolation method. The interpolation model was developed based on the fundamental polynomial regression from first to third-degree polynomials. The equations for each polynomial degree are presented in Equations (10), (11), and (12), respectively.

$$f(x) = a + b_1X \tag{10}$$

$$f(x) = a + b_1X + b_2X^2 \tag{11}$$

$$f(x) = a + b_1X + b_2X^2 + b_3X^3 \tag{12}$$

The $R^2$ and RMSE values were also considered when choosing the degree of the polynomial to use. Table 4 presents the evaluation of bio-yield stress data augmentation by polynomial interpolation method degrees one to three.

**Table 4.** Evaluation of augmentation bio-yield stress data.

| Model | $R^2$ | RMSE (kPa) |
|---|---|---|
| $f(x) = a + b_1X$ | 0.9987 | 7.8688 |
| $f(x) = a + b_1X + b_2X^2$ | 0.9991 | 6.6765 |
| $f(x) = a + b_1X + b_2X^2 + b_3X^3$ | 0.9986 | 8.6121 |

### 2.3.2 Multiple linear regression (MLR)

Linear regression is a common supervised learning technique. It determines the relationship between a numerical result variable and a predictor variable. The number of predictor variables is the only distinction between simple and multiple linear regression. Like basic linear regression, multiple linear regression employs the least squares method to modify coefficients. In this study, we used multiple linear regression to build the model. The independent variable is the parameter that determines the dependent variable is calculated. Multiple linear regression models use multiple explanatory variables. This study uses several independent variables in environmental data, such as temperature, relative humidity, and storage time. The dependent variables are bio-yield stress and weight loss (Stangierski, Weiss and Kaczmarek, 2019).

An MLR model was developed to predict the bio-yield stress and weight loss ($y$) of onions from a linear combination of time ($t$), temperature ($T$), and relative humidity ($RH$) data, as described in Equation () with coefficients $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$, and standard estimation error $\varepsilon$.

$$y = \beta_0 + \beta_1 t + \beta_2 T + \beta_3 RH + \varepsilon \quad (13)$$

### 2.3.3 Partial least square regression (PLSR)

Partial least square regression (PLSR) is a recently developed technique that generalizes and combines characteristics of principle component analysis and multiple regression. It is especially helpful for predicting a collection of dependent variables from an enormous number of independent factors. PLSR identifies X components that are also related to Y. PLSR particularly seeks a collection of components known as latent variables that conduct simultaneous decomposition of X and Y under the constraint that these components explain as much of the covariance between X and T as feasible. This phase broadens PCA. In the subsequent regression stage, the decomposition of X is utilized to predict Y (Farifteh et al., 2007).

PLSR is a unique approach for multivariate data analysis created through real-world application. This technique is mainly employed to model linear regression between multi-dependent and multi-independent variables. In addition, this approach provides additional benefits that conventional multiple linear regression does not. When the number of observations is fewer than the number of variables, it avoids the adverse effects of multicollinearity and regression in modeling. In addition, PLSR incorporates the fundamental functions of regression models. Moreover, the method of PLSR may simultaneously model multiple response variables while successfully dealing with highly correlated and noisy independent variables (Lee, Huh and Park, 2014).

The objective of the PLSR algorithm, described in Equation (14), is to extract meaningful components (factors) $\{t_i\}$ from X. These components are retrieved in descending order of covariance Cov-measured importance ($t_i$, $Y$). $T = XW$, where the columns of $W$ are the weight vectors for the $X$ columns (Cheng and Wu, 2006).

PLSR was model used in this study for prediction modeling based on the same dataset as in the MLR model. For the PLSR model, the number of latent variables was determined to decompose the input variables (time, temperature, and relative humidity) and the output variable (bio-yield strength and weight loss). The decomposition of the variables was performed in a way that the covariance between the input variables ($X$) and the output variable ($Y$) becomes it is largest using the score matrix $T$ and loading matrices $P$ and $Q$ (Equation (14)). $E$ and $F$ are the residual matrices of the PLSR model.

$$\begin{aligned} X &= TP^T + E \\ Y &= TQ^T + F \end{aligned} \tag{14}$$

where, $X$ is input data, $Y$ is the predicted values, $T$ is a scoring matrix, $P$ and $Q$ are loading matrix $E$ and $F$ are residual matrices

### 2.3.4 Support vector regression (SVR)

SVM, also known as Support Vector Machines, is one of the most popular and commonly used classification algorithms in machine learning, and their application has been extended to regression analysis. However, the use of SVMs for regression is poorly described. This algorithm recognizes the existence of nonlinearity in the data and generates an accurate prediction model. In Support Vector Regression, the needed straight line to match the data is known as the hyperplane (Parveen, Zaidi and Danish, 2017).

A support vector regression approach aims to identify the $f(x)$ with the most significant $\varepsilon$ deviation from the actual objectives achieved for all training data. Any divergence more significant than $\varepsilon$ is unacceptable (Liu *et al.*, 2013). In order to achieve the above objective, SVR considers the following linear estimating function:

$$f(x) = a\phi(x) + b \tag{15}$$

where, $a$ and $b$ are coefficients, $\phi(x)$ denotes the high dimensional feature space, which is nonlinearly mapped from the input space $x$.

A kernel is a collection of mathematical functions that receive data as input and turn it into the desired form. These are typically employed to locate a hyperplane in higher-dimensional space. Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid are the most used kernels. RBF is used as the kernel by default. Each of these kernels is utilized based on the dataset.

There are also boundaries line. These are the two lines drawn at a distance from the hyperplane $\varepsilon$ (epsilon). Figure 9 represents the support vector regression model. It is utilized to provide a space between the data points. Where the data points represent the predicted values ($y_{pre}$), and the line represents the actual data ($y_{act}$) data. The two dashed lines are the bounds that are $\varepsilon$ distance away from the reference data, where $\varepsilon$ is a parameter chosen by the user (Liu *et al.*, 2013).

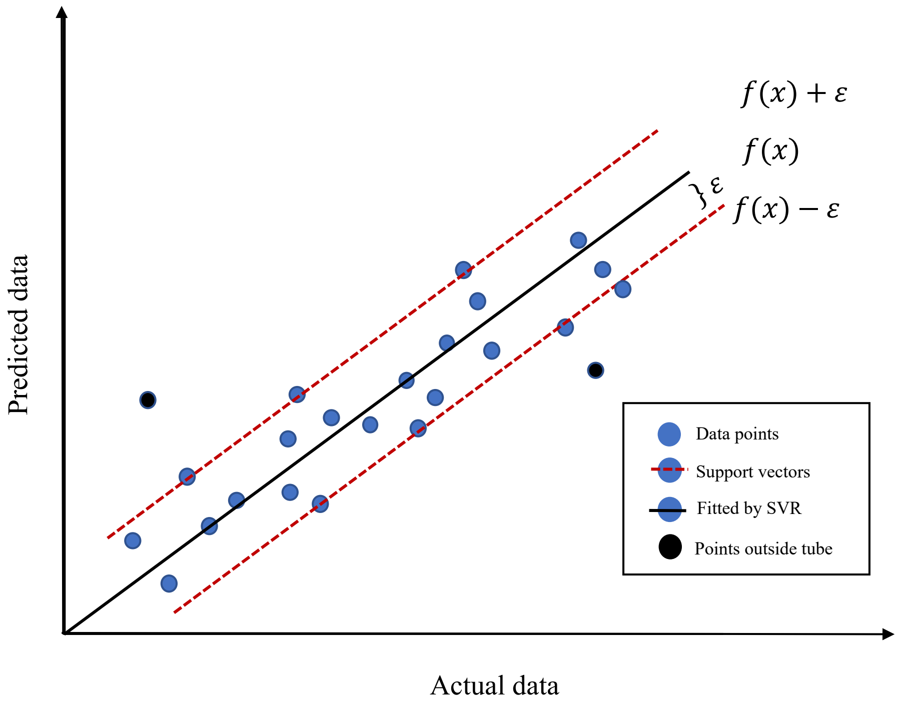



**Figure 9.** Graph of the support vector regression (SVR) model.

SVR requires user-defined arguments to configure kernel-specific settings. In addition, it is necessary to determine the ideal values of the legality argument C and the size errors in the sensitive region. The settings selection determines the complexity of the forecast. The dimensionality of the feature space is governed by the choice of kernel function and its parameters, while model complexity is determined by an additional penalty parameter.

The parameter ε represents the difference between actual values and regression function values. This distinction can be shown as a tube around the regression function. C defines a cost function that measures empirical risk; it represents a parameter that determines the trade-off between empirical risk and model flatness. The constant C > 0 represents the degree of penalty for a sample whose error exceeds $\varepsilon$ (Liu *et al.*, 2013).

In this study, three kinds of kernels were used, including linear, polynomial, and radial basis functions (RBF). Optimization of the parameters required for each kernel was performed through cross-validation. For the polynomial kernel, the degree of the kernel was determined. For the RBF kernel, regularization parameter ($C$), kernel coefficient gamma ($\gamma$), and epsilon ($\varepsilon$) were determined.

# Chapter 3. Result and Discussion

## 3.1 Environmental storage data

The most important thing in storing onions is the storage environment. The temperature and humidity affect the physicochemical properties of the onion bulb, which is essential to the degradation of the quality of the stored onions. Figure 10 shows the graph representing the weekly median of sensor data during the 10-week experiment period. The onions were stored at a temperature of 0~1°C. It was observed that the air surface temperature obtained with a temperature sensor between the onions was higher than the body temperature inserted inside them. It is presumably caused by the space between the onion bulbs, allowing gas exchange and heat transfer between the storage temperature and the temperature of the onion bulb. The humidity was not controlled, so there is no significant difference shown as shown in Figure 11, it remained above 90% throughout the experiment.

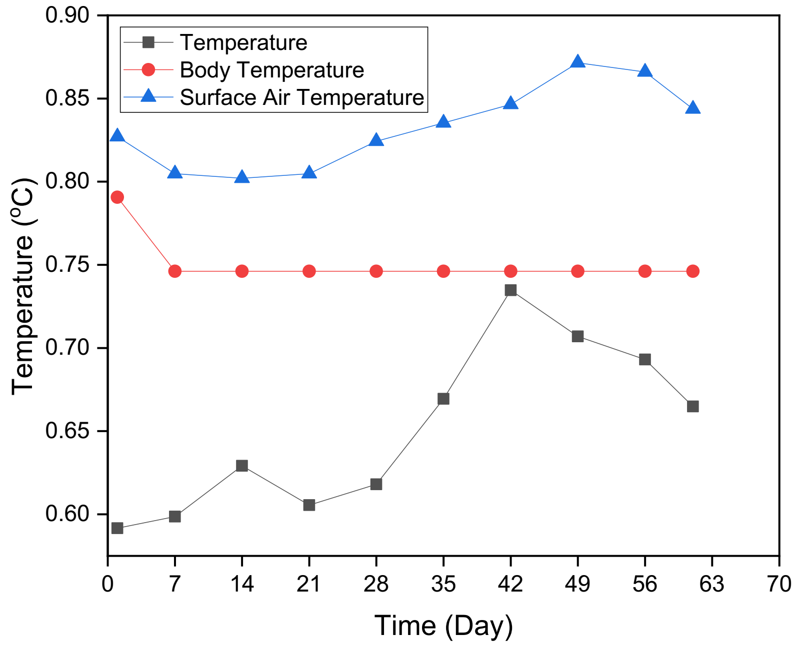



**Figure 10.** Temperature data for ten weeks storage periods.

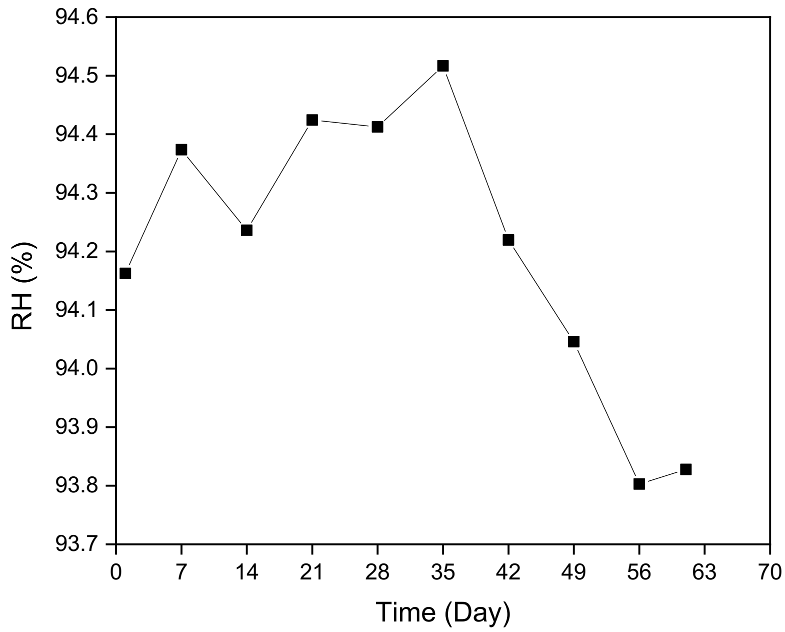



**Figure 11.** Relative humidity data for ten weeks storage periods.

## 3.2 Result of Biophysical measurement

### 3.2.1 Bio-yield stress

Experiment data for the compression test is shown graphically in Figure 12 for weeks 1, 4, 8, and 10. The force is shown to decrease with time. The deformation, though, is getting worse. As the force required to deform the onion increases, the graph suggests that the onion is becoming softer. The graphs demonstrate that the onions' quality changed as a function of storage conditions and periods.

Figure 13 shows the bio-yield stress values measured in a weekly experiment. Bio-yield stress is stress in the bio-yield point where the bio-yield point is related to failure in the material's microstructure associated with an initial cellular structure. Therefore, the reduction in bio-yield stress is an essential indicator of the chemical change in the onion being stored. The bio-yield stress at the beginning of storage time shows about 600 kPa to 800 kPa. Then it decreased over time, presumably because the active transpiration influenced water loss and a low-temperature treatment. As a result, it caused shrinking due to dehydration and softening due to chilling injury. Therefore, if the experiment had been continued after 10 weeks, the bio-yield stress value would have continued to decrease. Since curing was not performed before the experiment, the moisture content of onions may have been slightly higher than the average onions, and it caused the onion easier to decay.

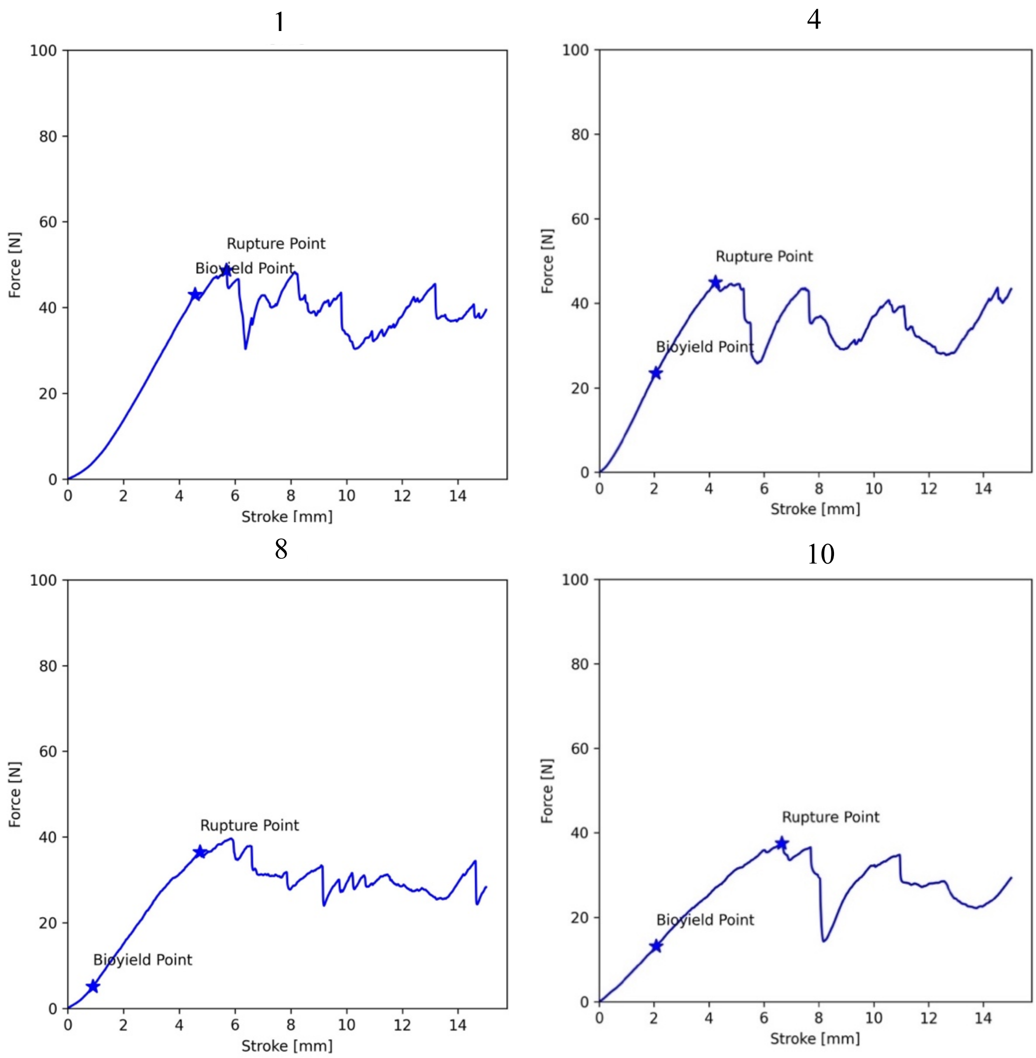



**Figure 12.** Graphs of compression test in weeks 1, 4, 8 and 10.

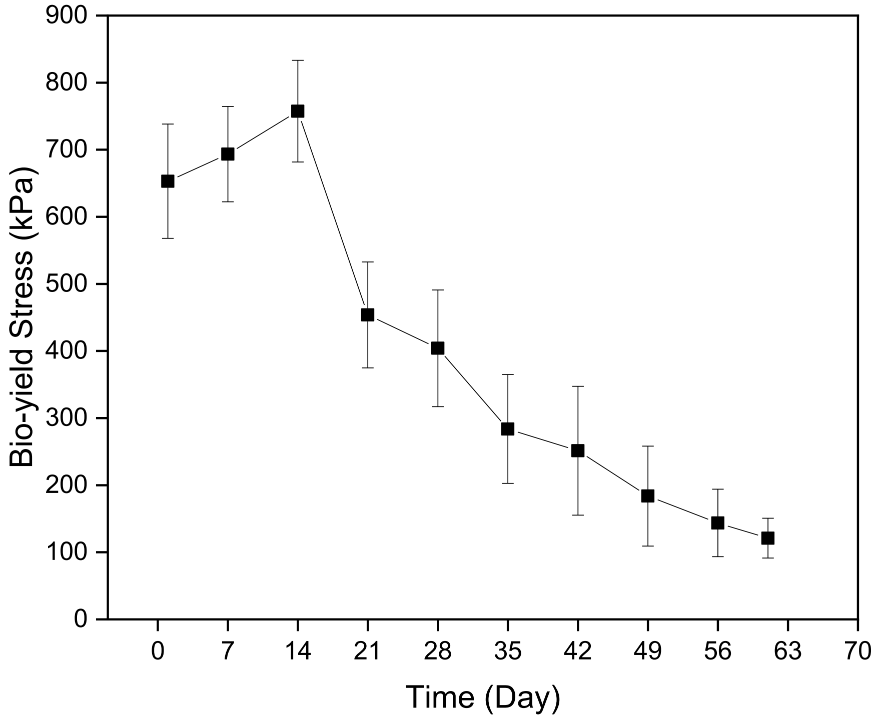



**Figure 13.** Bio-yield stress for ten weeks storage period. The data are expressed as mean ± SD of 20 samples. Vertical bars represent the standard errors of the means.

### 3.2.2 Respiration rate

Figure 14 shows the respiration rate of onions. Onions are non-climacteric type so, the respiration rate does not increase rapidly during aging. The respiration rate ranges from 150 to 250 ml $CO_2$ kg/h for ten weeks storage period. It is higher than any previous study that states the onion respiration rate. It is stated that onion respiration rate for $CO_2$ consumption is about 16 to 23.4 ml $CO_2$ kg/h in a low storage condition (WARD and TUCKER, 1976). The difference amount is presumably because the sample we used in this study is a non-curing sample where the moisture content has not been reduced after harvesting. As a result, the respiration rate shows a non-linearity to storage periods.

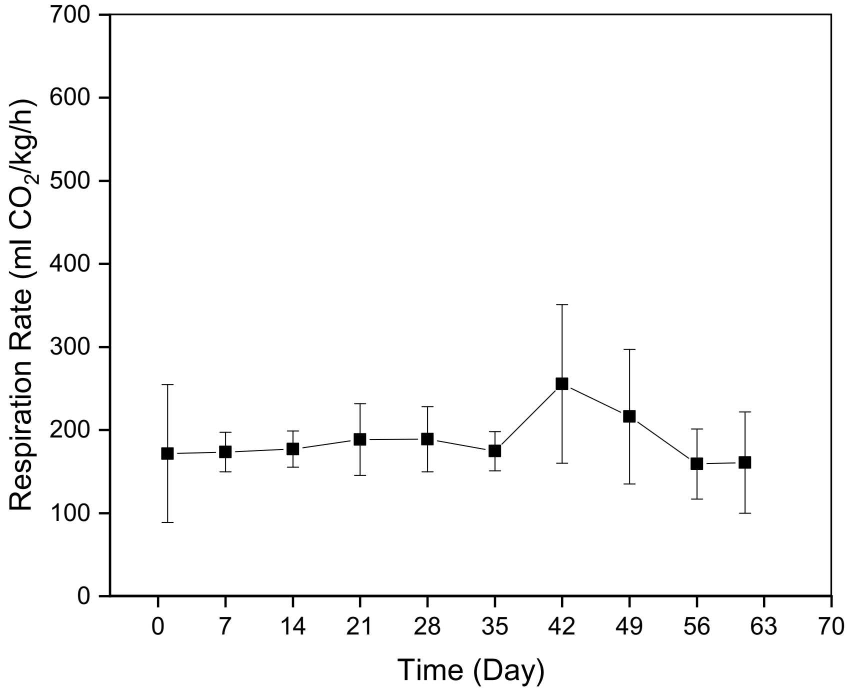



**Figure 14.** Respiration rate for ten weeks storage period. The data are expressed as mean ± SD of 5 samples. Vertical bars represent the standard errors of the means.

### 3.2.3 Weight loss

As shown in Figure 15, the weight of onions decreased by about 600g during the experiment period and dividing it by the total onion weight corresponds to 3.06 g/kg. Since it occurred for 10 weeks, the weight loss of about 0.306 g/kg occurred every week. Figure 16 shows how much weight loss occurred each week. And when compared to estimated weight loss, which is theoretically calculated from eq. (6), the estimated values are calculated based on the temperature and humidity. The contrast between the measured and estimated values is lightly discorded. Hypothetically, the reduced weight might be influenced by temperature and humidity. Comparing measured and estimated data suggests that transpiration is the most influential mechanism for weight loss. Therefore, weight loss can be a crucial parameter for evaluating the freshness and quality of objects. Figure 17 shows that the weight loss for 10 weeks of storage periods reached a total of >30 g and linearly correlated with the storage time.

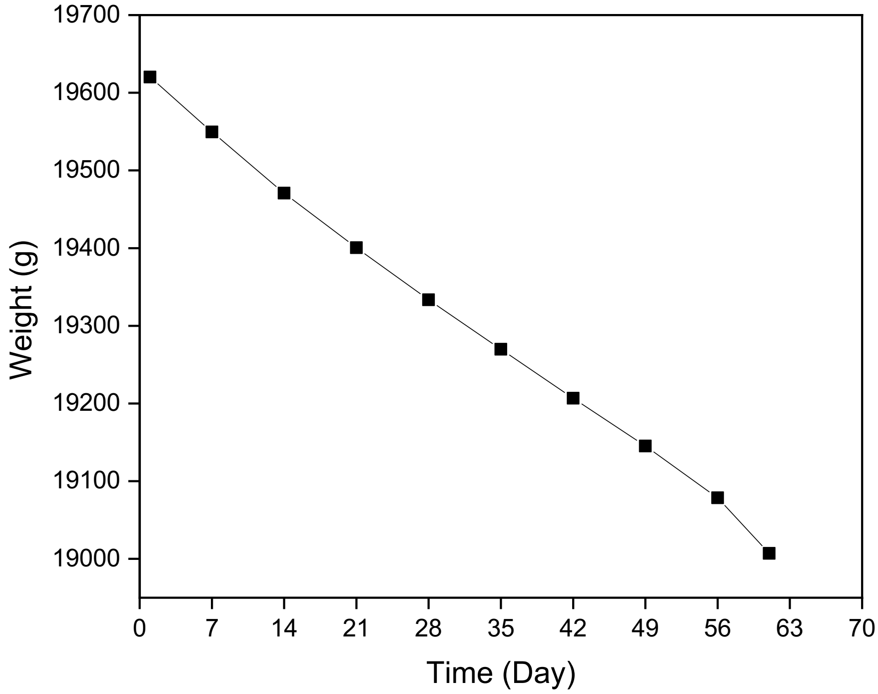



**Figure 15.** Weight data for ten weeks storage period.

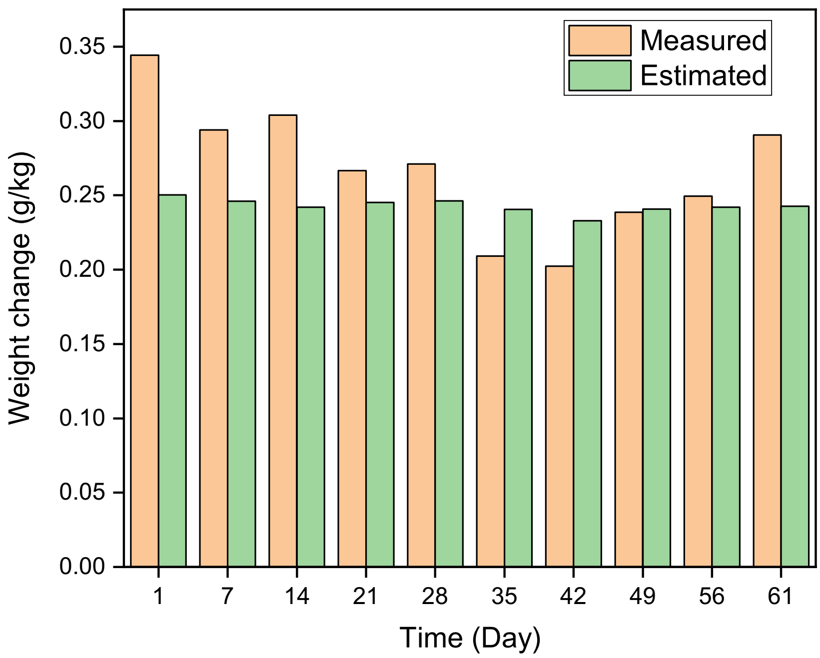



**Figure 16.** Weekly weight change during the storage period.

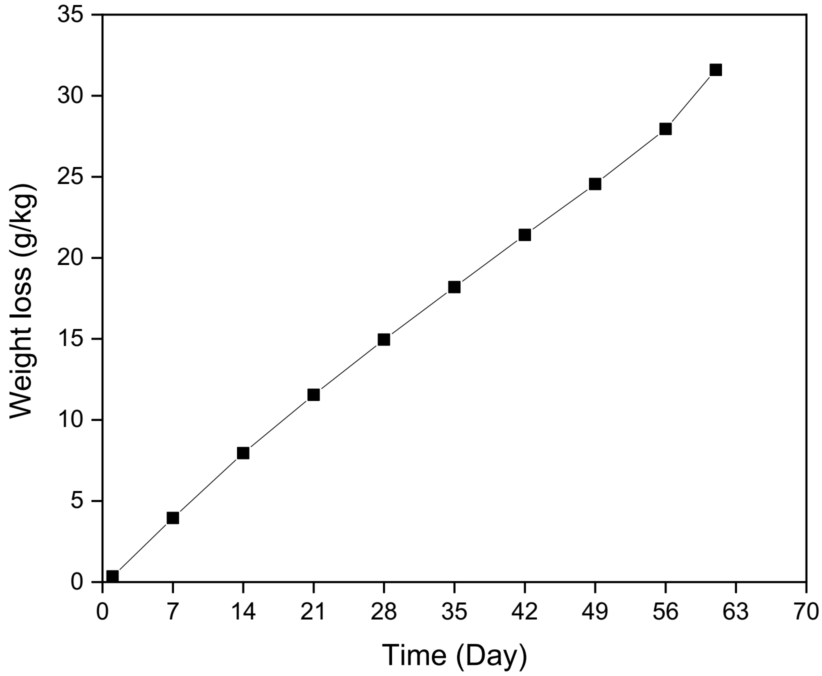



**Figure 17**. Weight loss for ten weeks storage period.

## 3.3 Prediction model evaluation

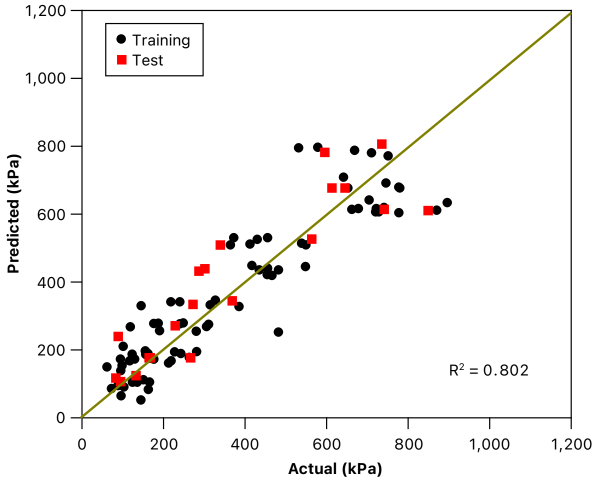
In this study, we measure bio-yield stress, weight loss, and respiration rate as the quality attribute to observe the quality change of onions during storage. However, in the case of respiration rate, it was impossible to develop a meaningful machine learning-based quality estimation model due to its nonlinearity. Considering this result, the quality estimation model of onion was performed only using bio-yield stress and weight loss parameters.

### 3.3.1 Bio-yield stress

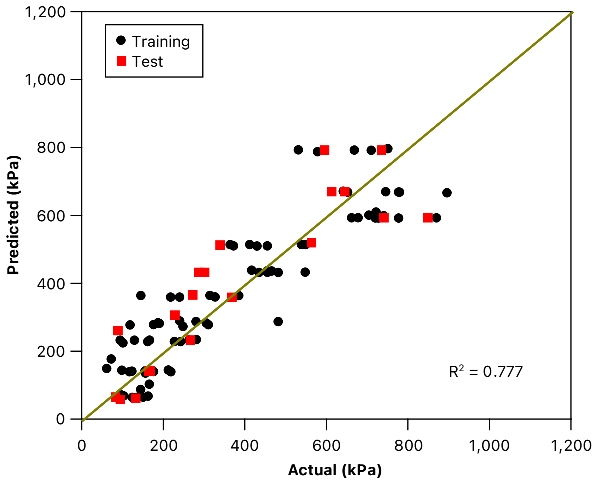
Table 5 presents evaluation metrics of regression models on train and test set. In PLSR, the highest performance was shown in cross-validation process when two latent variables were used. In SVR, a model with RBF kernel showed the highest performance, and $C$, γ, and ε were optimized to 0.0001, 10000, and 0.00001, respectively. $R^2$ of the models showed values of >0.8 for MLR and SVR, and 0.7773 for PLSR, the lowest among the three models. These results are interpreted as PLSR, which converts multivariate data into a small number of latent variables for regression, which is less suitable in our model, where only three variables are used as input. Similar values were shown in $R^2$ and RMSE of MLR and SVR, indicating that the optimized model can be obtained only with linear regression. In the case of MAPE, SVR showed a lower value than MLR, which is considered because the error of SVR appeared relatively low in the region where the values of the dependent variables were high. Figure 18 shows scatter plots of trained machine-learning models' actual and predicted bio-yield stress. Models generally showed similar trends, and the error increased in the high bio-yield stress range data.

**Table 5**. Evaluation of bio-yield stress estimation models.

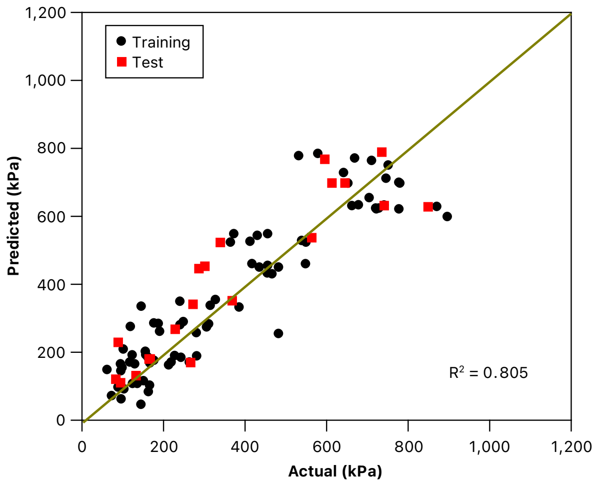
| Method | $R^2$ | | RMSE (kPa) | | MAPE (%) | |
|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test |
| MLR | 0.836 | 0.802 | 96.87 | 106.7 | 28.47 | 29.12 |
| PLSR | 0.822 | 0.777 | 101.3 | 113.4 | 32.78 | 33.87 |
| SVR | 0.837 | 0.805 | 96.41 | 106.1 | 28.68 | 25.86 |




(a)




(b)




(c)

**Figure 18.** Scatter plots of actual and predicted bio-yield stress: (a) MLR (b) PLSR (c) SVR.

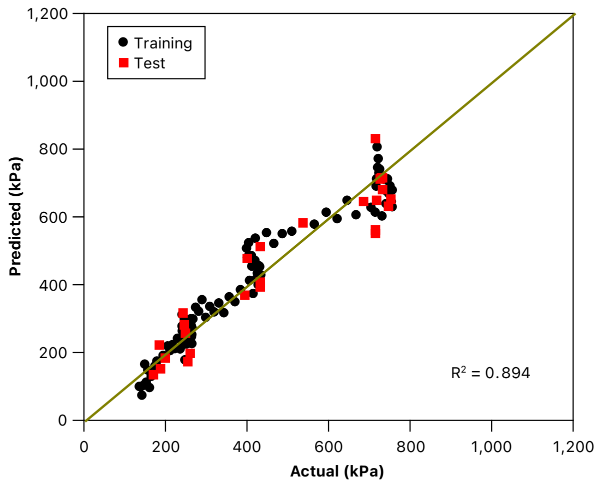
Along with an increasing degree of the polynomial, the complexity of the model also increases. Therefore, the n value must be chosen precisely. If this value is low, then the model will not be able to fit the data correctly. If high, the model will overfit the data easily. as described in Table 4, the degree interpolation method that is used in the following regression technique is a polynomial technique with degree 2 since it has the highest accuracy.

Table 6 presents evaluation metrics of regression models on train and test sets after data augmentation using polynomial degree 2. In PLSR, the highest performance was shown in the cross-validation process when two latent variables were used. In SVR, the model with RBF kernel showed the highest performance, and $C$, $\gamma$, and $\varepsilon$ were optimized to 0.0001, 10000, and 0.00001, respectively. $R^2$ of the models showed values of 0.88 for MLR and SVR and 0.87 for PLSR, the lowest among the three models. $R^2$ and RMSE of MLR and SVR showed a similar value. Figure 19 shows scatter plots of trained actual and predicted bio-yield stress using the augmentation bio-yield dataset. Models generally showed similar trends, and the error increased in the high bio-yield stress range data.

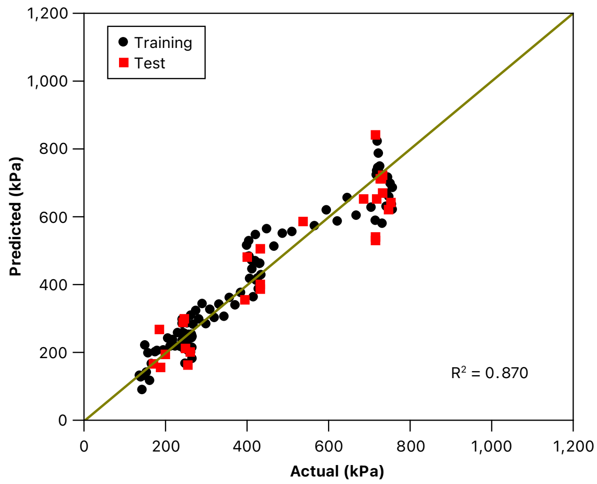
Compared to the regression models without augmentation data, the model performance in the augmentation dataset shows higher accuracy with $R^2$ values 0.88 for SVR and MLR and 0.89 and 0.87 for PLSR model, and the model of a non-augmentation database, 0.80, 0.7, and 0,81 for MLR, PLSR, and SVR, respectively. Even if it has a better result, we considered building the regression models utilizing a database without augmentation. Due to the fact that the interpolation approach may not correctly reflect the current values, despite the fact that it may be accurate for some specific numbers.

**Table 6.** Evaluation of bio-yield stress estimation models with augmentation data.

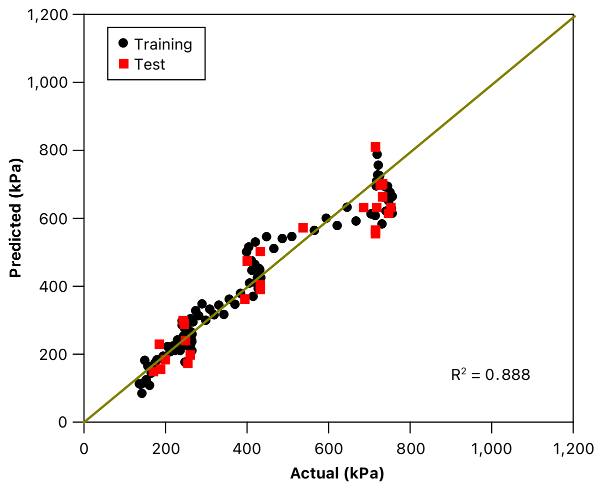
| **Method** | **$R^2$** | | **RMSE (kPa)** | | **MAPE (%)** | |
|---|---|---|---|---|---|---|
| | **Train** | **Test** | **Train** | **Test** | **Train** | **Test** |
| MLR | 0.937 | 0.894 | 50.07 | 72.58 | 0.106 | 0.138 |
| PLSR | 0.927 | 0.870 | 54.06 | 80.26 | 0.112 | 0.148 |
| SVR | 0.934 | 0.888 | 51.18 | 74.62 | 0.098 | 0.140 |




(a)




(b)




(c)

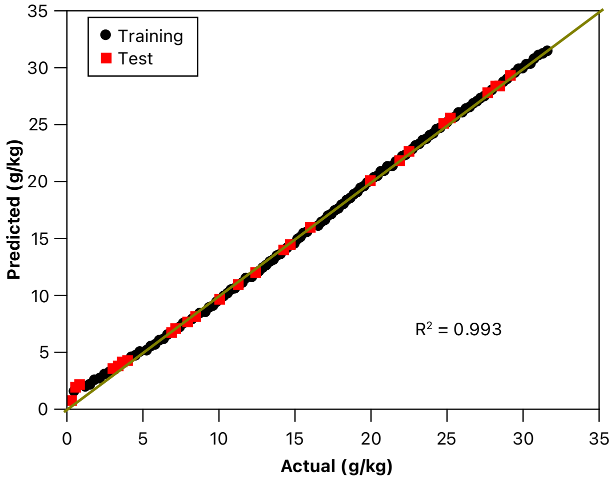
**Figure 19.** Scatter plots of actual and predicted bio-yield stress with augmentation data: (a) MLR (b) PLSR (c) SVR.

### 3.3.2 Weight loss rate

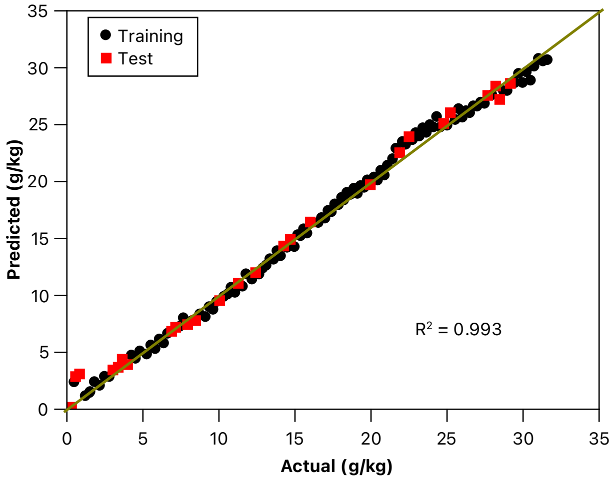
Table 7 presents evaluation metrics of regression models on the train and test set. In PLSR, the highest performance was shown in the cross-validation process when two latent variables were used. In SVR, a model with RBF kernel showed the highest performance, and $C$, γ, and ε were optimized to 1000, 0.001, and 0.1, respectively. $R^2$ of all the models showed values of >0.99 for MLR, PLSR, and SVR. This result is interpreted as the models having good accuracy in all models, which indicates that weight loss increased almost linearly with storage time. The result of RMSE values shows that PLSR has the highest error. It is about 0.83. 0.57 and 0.48 g/kg for SVR and MLR, respectively. In the case of MAPE, SVR shows a low value than MLR, which is considered because the error of SVR appeared relatively low in the region where the values of the dependent variables were high. Figure 20 shows scatter plots of trained machine-learning models' actual and predicted weight loss. Models generally showed similar trends, and the error increased in the high weight loss range data.

**Table 7.** Evaluation of weight loss estimation models.

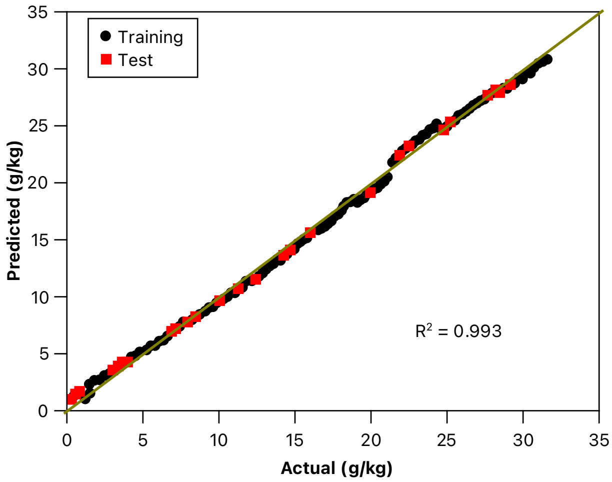
| Method | $R^2$ | | RMSE (g/kg) | | MAPE (%) | |
|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test |
| MLR | 0.999 | 0.997 | 0.315 | 0.482 | 0.065 | 0.234 |
| PLSR | 0.995 | 0.993 | 0.617 | 0.834 | 0.075 | 0.831 |
| SVR | 0.995 | 0.996 | 0.570 | 0.538 | 0.066 | 0.211 |




(a)




(b)




(c)

**Figure 20.** Scatter plots of actual and predicted weight loss: (a) MLR (b) PLSR (c) SVR

# Chapter 4. Conclusion

In the experiment of this study, environmental data, including temperature and relative humidity, and onion physical properties, including weight, bio-yield stress, and onion respiration rate, were measured. Three different models were built to predict onion quality using environmental data to predict bio-yield stress and weight loss during storage. For bio-yield stress, PLSR was trained using 2 latent variables to build the models, and SVR was trained using the best hyperparameters of $C$, γ, and ε and RBF kernel. The models were evaluated by calculating the evaluation metrics, including $R^2$, RMSE, and MAPE. In the measurement data, the bio-yield stress showed a decreasing trend. The storage temperature was relatively stable during 10 weeks of storage in 0-2℃, and the humidity during storage remained above 90%. On the other hand, the weight loss was increased along with storage periods, and the respiration rate remained same in the range 150-250 ml $CO_2$ kg/h. For the prediction of the bio-yield stress of onion, we found that SVR and MLR could be used to predict the bio-yield stress quality of onion during storage with $R^2$ values of >0.8.

The same process was used in building the prediction model for weight loss. Environmental data, including temperature, relative humidity, and time used as input data, and weight loss was measured. The weight loss is increased over time, and for the weight loss prediction model, the $R^2$ values of MLR, PLSR and SVR are >0.99. where it shows excellent accuracy. The input data: time, temperature, and relative humidity are almost linearly with the weight loss.

On the contrary, it showed significant nonlinearity in the respiration rate, and it was impossible to develop a quality estimation model by applying machine learning techniques. It was judged to be a task that required advanced techniques and made it difficult to obtain meaningful results with the current amount of data.

This study presents quality change results during the storage of onions and models for predicting onion quality. Although the modeling results showed a correlation between environmental data and onion quality, there is a limitation in that the experiment was conducted for one condition, and the data from one experimental batch was used for modeling. Since the correlation between data exists within one batch of time series data, it is considered that additional experiments with various storage conditions and appropriate dataset splits are required in further research for the training and evaluation of robust models.

# References


Abbott, J.A. (1999) ‘Quality measurement of fruits and vegetables’, *Postharvest Biology and Technology*, 15(3), pp. 207–225. Available at: https://doi.org/10.1016/S0925-5214(98)00086-6.

Accorsi, R. *et al.* (2022) ‘Simulating product-packaging conditions under environmental stresses in a food supply chain cyber-physical twin’, *Journal of Food Engineering*, 320(August 2021), p. 110930. Available at: https://doi.org/10.1016/j.jfoodeng.2021.110930.

ASAE, standard 368. 4 (DEC 2000) (2008) ‘Compression test of food materials of convex shape’, *American Society of Agricultural and Biological Engineers*, 2000(MAR95), pp. 580–587. Available at: http://elibrary.asabe.org/abstract.asp?aid=42544&t=2.

Aung, M.M. and Chang, Y.S. (2014a) ‘Temperature management for the quality assurance of a perishable food supply chain’, *Food Control*, 40(1), pp. 198–207. Available at: https://doi.org/10.1016/j.foodcont.2013.11.016.

Aung, M.M. and Chang, Y.S. (2014b) ‘Traceability in a food supply chain: Safety and quality perspectives’, *Food Control*, 39(1), pp. 172–184. Available at: https://doi.org/10.1016/j.foodcont.2013.11.007.

Badia-Melis, R., Mishra, P. and Ruiz-García, L. (2015) ‘Food traceability: New trends and recent advances. A review’, *Food Control*, 57, pp. 393–401. Available at: https://doi.org/10.1016/j.foodcont.2015.05.005.

BAEK, H.-S. and KIM, I.-S. (2020) ‘An Analysis of the Impact of Climate Change on the Korean Onion Market’, *Journal of Industrial Distribution & Business*, 11(3), pp. 39–50. Available at: https://doi.org/10.13106/jidb.2020.vol11.no3.39.

Bovi, G.G. *et al.* (2016) ‘Transpiration and moisture evolution in packaged fresh horticultural produce and the role of integrated mathematical models: A review’, *Biosystems Engineering*, 150, pp. 24–39. Available at: https://doi.org/10.1016/j.biosystemseng.2016.07.013.

Brusewitz, G.H., McCollum, T.G. and Zhang, X. (1991) 'Impact bruise resistance of peaches', *Transactions of the American Society of Agricultural Engineers*, 34(3), pp. 962–965. Available at: https://doi.org/10.13031/2013.31756.

Chen, C.R., Ramaswamy, H.S. and Alli, I. (2001) 'Prediction of quality changes during osmo-convective drying of blueberries using neural network models for process optimization', *Drying Technology*, 19(3–4), pp. 507–523. Available at: https://doi.org/10.1081/DRT-100103931.

Chen, R.Y. (2017) 'An intelligent value stream-based approach to collaboration of food traceability cyber physical system by fog computing', *Food Control*, 71, pp. 124–136. Available at: https://doi.org/10.1016/j.foodcont.2016.06.042.

Cheng, B. and Wu, X. (2006) *An Modified PLSR Method in Prediction*, *Journal of Data Science*.

Cho, J.-E., Bae, R.-N. and Lee, S.-K. (2010) 'Current Research Status of Postharvest Technology of Onion (Allium cepa L.)', *Horticultural Science & Technology*, 28(3), pp. 522–527.

Correa, E.C. *et al.* (2014) 'Advanced Characterisation of a Coffee Fermenting Tank by Multi-distributed Wireless Sensors: Spatial Interpolation and Phase Space Graphs', *Food and Bioprocess Technology*, 7(11), pp. 3166–3174. Available at: https://doi.org/10.1007/s11947-014-1328-4.

Correa-mosquera, A.R., Quicaz, M.C. and Zuluaga-domínguez, C.M. (2022) 'Shelf-life prediction of pot-honey subjected to thermal treatments based on quality attributes at accelerated storage conditions', *Food Control*, 142(July). Available at: https://doi.org/10.1016/j.foodcont.2022.109237.

Devahastin, S. and Niamnuy, C. (2010) 'Modelling quality changes of fruits and vegetables during drying: A review', *International Journal of Food Science and Technology*, pp. 1755–1767. Available at: https://doi.org/10.1111/j.1365-

2621.2010.02352.x.

Eboibi, O. and Uguru, H. (2017) ‘Storage conditions effect on physic-mechanical properties of Nandini cucumber’, *International Journal of Engineering and Technical Research (IJETR)*, 7(11), pp. 48–56.

Emana, B. *et al.* (2017) ‘Assessment of postharvest losses and marketing of onion in Ethiopia’, *International Journal of Postharvest Technology and Innovation*, 5(4), pp. 300–319. Available at: https://doi.org/10.1504/IJPTI.2017.092466.

Escobedo-Avellaneda, Z. *et al.* (2012) ‘Inclusion of the variability of model parameters on shelf-life estimations for low and intermediate moisture vegetables’, *LWT*, 47(2), pp. 364–370. Available at: https://doi.org/10.1016/j.lwt.2012.01.032.

Falayi, F.R., Yusuf, H.A. and State, O. (2014) ‘Performance Evaluation of a Modified Onion Storage Structure’, *journal of emerging trends in Engineering and Applied Science*, 5(6), pp. 334–339.

Farifteh, J. *et al.* (2007) ‘Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN)’, *Remote Sensing of Environment*, 110(1), pp. 59–78. Available at: https://doi.org/10.1016/j.rse.2007.02.005.

Ferreira, A.P.S. *et al.* (2015) ‘Storage of onions in farm scale ventilated silos’, *Acta Horticulturae*, 1099(November), pp. 123–128. Available at: https://doi.org/10.17660/ActaHortic.2015.1099.11.

Hartley, J. *et al.* (2018) ‘Wireless Sensor Networks for Healthcare’, 98(11).

Huang, X. *et al.* (2020) ‘Shelf-life Prediction of Chilled Penaeus vannamei Using Grey Relational Analysis and Support Vector Regression’, *Journal of Aquatic Food Product Technology*, 29(6), pp. 507–519. Available at: https://doi.org/10.1080/10498850.2020.1766616.

Isma’ila M, E. Karu, D. A Zhígila, Y.U.A. (2017) ‘Postharvest Storage and Shelf Life Potentials among Selected Varieties of Onion ( Allium cepa L . )’, *Scholars*

*Academic Journal of Biosciences*, 5(4), pp. 271–277. Available at: https://doi.org/10.21276/sajb.

Jang, S.-H. and Lee, S.-K. (2009) ‘Current Research Status of Postharvest Technology of Onion’, *Korean Journal of Horticultural Science and Technology*, 27(3), pp. 511–520.

Kaymak-Ertekin, F. and Gedik, A. (2005) ‘Kinetic modelling of quality deterioration in onions during drying and storage’, *Journal of Food Engineering*, 68(4), pp. 443–453. Available at: https://doi.org/10.1016/j.jfoodeng.2004.06.022.

Lee, B.J. *et al.* (2018) ‘Discrimination and prediction of the origin of Chinese and Korean soybeans using Fourier transform infrared spectrometry (FT-IR) with multivariate statistical analysis’, *PLOS ONE*, 13(4), pp. 1–16. Available at: https://doi.org/10.1371/journal.pone.0196315.

Lee, S.-Y., Huh, M.-H. and Park, M. (2014) ‘A modified partial least squares regression for the analysis of gene expression data with survival information’, *Journal of the Korean Data and Information Science Society*, 25(5), pp. 1151–1160. Available at: https://doi.org/10.7465/jkdi.2014.25.5.1151.

Lim, J. *et al.* (2014) ‘Non-destructive and Rapid Prediction of Moisture Content in Red Pepper (Capsicum annuum L.) Powder Using Near-infrared Spectroscopy and a Partial Least Squares Regression Model’, *Journal of Biosystems Engineering*, 39(3), pp. 184–193. Available at: https://doi.org/10.5307/jbe.2014.39.3.184.

Liu, S. *et al.* (2013) ‘A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction’, *Mathematical and Computer Modelling*, 58(3–4), pp. 458–465. Available at: https://doi.org/10.1016/j.mcm.2011.11.021.

Mallik, A. *et al.* (2018) ‘Monitoring food storage humidity and temperature data using IoT’, *MOJ Food Processing & Technology*, 6(4). Available at: https://doi.org/10.15406/mojfpt.2018.06.00194.

Masoudi, H., Tabatabaeefar, A. and Borghaee, A.M. (2007) ‘Determination of storage effect on mechanical properties of apples using the uniaxial compression test', *Canadian Biosystems Engineering / Le Genie des biosystems au Canada*, 49(January).

Melhem, M. *et al.* (2016) 'Regression Methods for Predicting the Product's Quality in the Semiconductor Manufacturing Process', *IFAC-PapersOnLine*, 49(12), pp. 83–88. Available at: https://doi.org/10.1016/j.ifacol.2016.07.554.

Mitra, J., Shrivastava, S.L. and Rao, P.S. (2015) 'Non-enzymatic browning and flavour Kinetics of vacuum dried onion slices', *International Agrophysics*, 29(1), pp. 91–100. Available at: https://doi.org/10.1515/intag-2015-0010.

Mohsenin, N.N. (2020) *Physical Properties of Plant and Animal Materials: v. 1: Physical Characteristics and Mechanical Properties*, *New York*. Available at: https://doi.org/10.4324/9781003062325.

Parveen, N., Zaidi, S. and Danish, M. (2017) 'Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr(VI)', *Process Safety and Environmental Protection*, 107, pp. 428–437. Available at: https://doi.org/10.1016/j.psep.2017.03.007.

Porras-Amores, C., Mazarrón, F.R. and Cañas, I. (2014) 'Study of the vertical distribution of air temperature in warehouses', *Energies*, 7(3), pp. 1193–1206. Available at: https://doi.org/10.3390/en7031193.

Ramzi, M. *et al.* (2015) 'Modeling of rheological behavior of honey using genetic algorithm-artificial neural network and adaptive neuro-fuzzy inference system', *Food Bioscience*, 9(1), pp. 60–67. Available at: https://doi.org/10.1016/j.fbio.2014.12.001.

Rozov, S. (2020) 'Machine Learning and Deep Learning methods for predictive modelling from Raman spectra in bioprocessing', (February).

Sanaeifar, A., Bakhshipour, A. and de La Guardia, M. (2016) 'Prediction of banana quality indices from color features using support vector regression', *Talanta*, 148, pp. 54–61. Available at: https://doi.org/10.1016/j.talanta.2015.10.073.

Sang, M.K. *et al.* (2014) 'Penicillium brasilianum as a novel pathogen of onion (Allium cepa L.) and other fungi predominant on market onion in Korea', *Crop Protection*, 65(August 2013), pp. 138–142. Available at: https://doi.org/10.1016/j.cropro.2014.07.016.

Sarmah, B. and Aruna, G. (2020) 'Detection of Food Quality and Quantity at Cold Storage using IoT', *2020 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2020*, pp. 200–203. Available at: https://doi.org/10.1109/WiSPNET48689.2020.9198348.

Sasongko, S.B. *et al.* (2020) 'Effects of drying temperature and relative humidity on the quality of dried onion slice', *Heliyon*, 6(7), p. e04338. Available at: https://doi.org/10.1016/j.heliyon.2020.e04338.

Sastry, S.K. (1985) 'Moisture losses from perishable commodities: recent research and developments', *International Journal of Refrigeration*, 8(6), pp. 343–346. Available at: https://doi.org/10.1016/0140-7007(85)90029-5.

Shao, P. *et al.* (2021) 'An overview of intelligent freshness indicator packaging for food quality and safety monitoring', *Trends in Food Science and Technology*, 118(PA), pp. 285–296. Available at: https://doi.org/10.1016/j.tifs.2021.10.012.

Sharma, K. *et al.* (2015) 'A comparative study of anaerobic and aerobic decomposition of quercetin glucosides and sugars in onion at an ambient temperature', *Frontiers in Life Science*, 8(2), pp. 117–123. Available at: https://doi.org/10.1080/21553769.2014.998298.

Snowden, A.L. (1992) *Post-Harvest Diseases and Disorders of Fruits and Vegetables: Volume 2: Vegetables*. 1st edn. CRC Press. Available at: https://doi.org/https://doi.org/10.1201/b18215.

Soliman, S.N. and El-Sayed, A.E. (2017) 'Penetration and Stress-Strain Behavior of Potato Tubers During Storage', *Misr Journal of Agricultural Engineering*, 34(4), pp. 2291–2310. Available at: https://doi.org/10.21608/mjae.2017.97514.

Stangierski, J., Weiss, D. and Kaczmarek, A. (2019) 'Multiple regression models and Artificial Neural Network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed Gouda cheese', *European Food Research and Technology*, 245(11), pp. 2539–2547. Available at: https://doi.org/10.1007/s00217-019-03369-y.

Vapnik, V.N. (2000) *The Nature of Statistical Learning Theory*. 2nd edn, *The Nature of Statistical Learning Theory*. 2nd edn. Edited by M. Jordan et al. New York: Springer. Available at: https://doi.org/10.1007/978-1-4757-3264-1.

Vázquez-Barrios, M.E. *et al.* (2006) 'Study and prediction of quality changes in garlic cv. Perla (Allium sativum L.) stored at different temperatures', *Scientia Horticulturae*, 108(2), pp. 127–132. Available at: https://doi.org/10.1016/j.scienta.2006.01.013.

WARD, C.M. (1976) 'The influence of temperature on weight loss from stored onion bulbs due to desiccation, respiration and sprouting', *Annals of Applied Biology*, 83(1), pp. 149–155. Available at: https://doi.org/10.1111/j.1744-7348.1976.tb01703.x.

WARD, C.M. and TUCKER, W.G. (1976) 'Respiration of maleic hydrazide treated and untreated onion bulbs during storage', *Annals of Applied Biology*, 82(1), pp. 135–141. Available at: https://doi.org/10.1111/j.1744-7348.1976.tb01680.x.

Xiao, X. *et al.* (2017) 'Improving traceability and transparency of table grapes cold chain logistics by integrating WSN and correlation analysis', *Food Control*, 73, pp. 1556–1563. Available at: https://doi.org/10.1016/j.foodcont.2016.11.019.

Xu, G., Shen, W. and Wang, X. (2014) 'Applications of wireless sensor networks in marine environment monitoring: A survey', *Sensors (Switzerland)*, 14(9), pp. 16932–16954. Available at: https://doi.org/10.3390/s140916932.

## Abstract in Korean

국문 초록

# 머신 러닝 기법을 이용한 양파 저장기간 중 품질 평가에 관한 연구 국문 초록

서울대학교 대학원

바이오시스템 •소개학부 바이오시스템공학전공

누르히스나 (Nandita Irsaulul Nurhisna)

양파는 한국의 주요 채소이다. 따라서 연중 꾸준히 발생하는 수요를 수용하기 위해서는 장기 보관이 필요하다. 그러므로, 양파의 저장 수명을 늘리기 위해서는 양파를 저장하는 방법에 유의해야 한다. 보관 중 온도와 상대습도는 양파의 품질을 변화시키는 중요한 역할을 하므로 온도와 습도 조절은 양파의 품질을 유지하기 위해 반드시 이루어져야 한다. 기계적 특성과 호흡 속도는 저장 기간 동안 관찰되는 양파의 주요 품질 특성으로 선택되었다. 또한 기존의 연구들이 저장된 양파의 품질 변화를 예측하기 위해 화학 역학을 기반으로 하는 모델만 사용했다는 점을 고려할 때, 기계학습을 활용하여 예측모델을 개발하는 방법은 충분히 고려될 만 하다.

본 연구에서는 2022년 3월부터 6월까지 10주간의 저장 기간 동안 양파를 0-1°C로 저장하면서 30분마다 환경 데이터를 수집하였으며, 매주 1회 파괴 실

험을 수행하였다. 저장고 내부에 설치된 센서에서 수집된 생체중 데이터와 환경 데이터(시간, 온도, 상대습도)를 이용하여 양파의 생물체 항복 강도 및 생체중 감소에 대한 예측 모델을 생성하기 위해 3가지 기계학습 기법을 사용하였다. 결과적으로 생물체 항복 강도 데이터 100개와 2차 다항식 보간법을 사용한 127개의 환경 데이터가 모델 개발에 사용되었다. 본 연구에서 사용된 기계 학습 기법은 다중 선형 회귀(MLR), 부분 제곱 최소 회귀(PLSR), 서포트 벡터 머신(SVR)이다. 데이터는 80:20의 비율로 트레이닝세트와 테스트세트로 나뉘었고, 트레이닝 세트의 학습 과정에서 10배 교차 검증이 수행되었다. 회귀 모델의 평가 기준으로는 결정 계수($R^2$), 평균 제곱근 오차(RMSE) 및 평균 절대 백분율 오차(MAPE)를 사용하였다.

데이터 수집 결과 양파의 생물체 항복 강도는 시간이 지날수록 감소하였으며 비급등형 호흡을 하는 양파의 특성상 호흡수는 시간에 관계없이 유지되는 경향을 보였기 때문에 생체중은 선형적으로 감소하였다. 기계학습 모델 개발 결과, MVR 및 SLR 모델을 사용하여 저장 중 양파의 품질 특성을 예측할 수 있었으며, 생물체 항복 강도를 예측한 결과의 경우 $R^2$ 값이 >0.8, 생체중 감소량 예측 모델은 $R^2$>0.99의 결과를 얻었다.

**주요어:** 그물 포장 양파, 무선 센서 네트워크, 품질 추정, 기계 학습 기술, 기계적 특성.

**학번:** 2020-20878