International Journal of Advanced Smart Convergence Vol.9 No.1 132-140 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.1.132

IJASC 20-1-16

A Prediction of Nutrition Water for Strawberry Production using Linear Regression

Saravanakumar Venkatesan[1], Sathishkumar V E[1], Jangwoo Park[2], Changsun Shin[2], Yongyun Cho[2,*]

[1]Ph.D. Candidate Student, Department of Information and Communication Engineering, Sunchon National University, Suncheon, South Korea.
[2]Professor, Department of Information and Communication Engineering, Sunchon National University, Suncheon, South Korea.
*Skumarvsk1288@gmail.com, sshb56@s.scnu.ac.kr, jwpark@scnu.ac.kr, csshin@scnu.ac.k, *yycho@scnu.ac.kr*

Abstract

It is very important to use appropriate nutrition water for crop growth in hydroponic farming facilities. However, in many cases, the supply of nutrition water is not designed with a precise plan, but is performed in a conventional manner. We proposes a forecasting technique for nutrition water requirements based on a data analysis for optimal strawberry production. To do this, the proposed forecasting technique uses linear regression for correlating strawberry production, soil condition, and environmental parameters with nutrition water demand for the actual two-stage strawberry production soil. Also, it includes predicting the optimal amount of nutrition water requires according to the heterogeneous cultivation environment and variety by comparing the amount of nutrition water needed for the growth and production of different kinds of strawberries. We suggested study uses two types of section beds that are compared to find out the best section bed production of strawberry growth. The dataset includes 233 samples collected from a real strawberry greenhouse, and the four predicted variables consist of the total amounts of nutrition water, average temperature, humidity, and CO_2 in the greenhouse.

Keywords: *Strawberry Growth, Linear Regression (LR), Correlation, IT-Agriculture, Prediction Model.*

1. Introduction

Strawberries are one of the most familiar fruits worldwide due to its gross value of production and nutrients. The growth of strawberries is considered as a full-fledged economy. Production in the greenhouse is increasing as domestic demand for strawberries increases for each season. Competitive strawberry growing in certain areas requires the knowledge of water requirements and physical or agricultural responses in potential constraints to maintain environmental sustainability.

This research used two types of strawberry bed sections to find out which bed best initiates the growth of

Manuscript Received: February. 2, 2020 / Revised: February. 6, 2020 / Accepted: February. 12, 2020
Corresponding Author: yycho@scnu.ac.kr
Tel: +82-61-750-3599, Fax: +82-61-750-3590
Department of Information and Communication Engineering, Sunchon National University, South Korea.

strawberries and produces with the help of linear regression and correlation method. Finally, in the results and discussion, we will show the best provision of Strawberry development. We manage two beds such as linear regression and correlation, which examine the relationship between the basic principles of strawberry growth and overall performance on strawberry growth. In summary, we also investigated whether the amount of water has any effect on change of proficiency.

2. Related works

We prove, we discuss the agriculture management and strawberry fruit production services described in detail, identifying related research works. The present forecasting results might help policy makers to develop macro-level policies for food security and more effective strategies for better planning strawberry production at the field scale can be generalized, analyzed this artificial forecasting harvest area and production of strawberry using time series analyses [1]. Application of precise control of apple scab and consequently significant reduction of the use of pesticides in apple production with benefits for environment, human health and economics of production was analyzed in the prediction of the apple scab using machine learning and simple weather stations [2]. For strawberry cultivated in closed solar greenhouse, we develop the machine learning models for forecasting agricultural products agricultural prediction model of transpiration rate of strawberry in closed cultivation based on the DBN-LS SVM (Deep Belief Network and Least Squares Support Vector Machine) algorithm [3]. The multispectral imaging technology, which is suggested together with a suitable analysis model, estimates rapidly quality attributes and classifies efficiently ripeness stages of strawberry, were analyzed the application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit [4].

The texture analysis for identifying roughness and ripeness stage of strawberry will be one of the scopes of this work. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry [5]. Because the forecast model effect was steady and the prediction method was also fast and easy, it became a new method for nitrate examination in water quality. Using research on spectral detection of nitrate in water quality based on K-S (Kennard-Stone) algorithm [6]. The aim of this study was to understand the flavor components of eating quality of several strawberries as a tool to develop were analyzed the a sensory and chemical analysis of fresh strawberries over harvest dates and seasons reveals factors that affect eating quality [7]. The SSC (Soluble Solid Content) was lower in fruit grown at 22°C (5.4 vs. 6.6), but there was no effect of transplant date. These results indicate that rising temperatures are responsible for the decline in soluble solid content of fruit at the end of the growing season in were analyzed the late season decline in strawberry fruit soluble solid content observed in Florida is caused by rising temperatures [8]. Strawberry breeding programs should be able to rely on valid and simple methodologies for evaluating sensory quality of new cultivars of the sensory characteristics of strawberry cultivars throughout the harvest season using projective mapping [9].

Artificial neural networks, fuzzy inference systems and the combination of these two are employed to develop the prediction models of output energies for broiler production using linear regression, ANN (Artificial Neural Network), MLP (Multi Layer Perceptron), RBF (Radial Basis Function), and ANFIS (Adaptive Neuro Fuzzy Inference System) models [10]. There is an analysis method research based on bigdata and artificial intelligence in order to distinguish fake news [11]. In this paper, we propose an ensemble model and apply it to classification problems. In diabetic pima indian's iris (immune reconstitution inflammatory syndrome) and semiconductor fault detection problem, the model proposed in [12] has a good performance in classification, compared to traditional single classifiers as like logistic regression, SVM (Support Vector Machine) and RF (Random Forest).

3. Materials and methods

In this study, we have used the Greenhouse strawberry data in the year of September 2015 to May 2016. They base these strawberry data on nutrition water, average temperature, humidity, and CO_2, etc. They collected 243 data from a local strawberry park in Korea. The statistical collection for the greenhouse strawberry growth was most popular among the past researchers, including linear regression. Correlation is the first step to data exploration before going into more deep analysis and to select the variable that might be related to one another, then the next step is to model the variable relationship using the linear regression that puts the relation between the response variable and the predicted variable.

Correlation, the persistence of correlation analysis is to measure and interpret the strength of a linear or nonlinear relationship between two continuous variables. When conducting correlation analysis, we use the term relationship to mean "linear association" (1,2). Herein, we focus on the Pearson and Spearman correlation coefficients. Both correlation coefficients take on values between -1 and +1, ranging from being negatively correlated (-1) to uncorrelated (0) to positively correlated (+1). The sign of the correlation coefficient (positive or negative) defines the direction of the relationship. The positive value indicates the strength of the correlation. We elaborate on two correlation coefficients, linear and rank, that are commonly used for measuring linear and general relationships between two variables.

In a linear regression, the independent variable to be predicted is identified using the independent variable, which is used for prediction. The data can both be predicted through user based or context-based after finding the correlation result.

3.1 Materials details

Fig. 1 shows the real data, which consist of strawberry nutrition water quantity, temperature, humanity and CO_2 sensed from the bed A of the strawberry greenhouse for 43 days. The changes of the nutrition water quantity values in A bed is 88.41 to 320.33 litter and production cell value is 0.88 to 17 kg, while bed nutrition water and production growth also proceeded for 43 days.

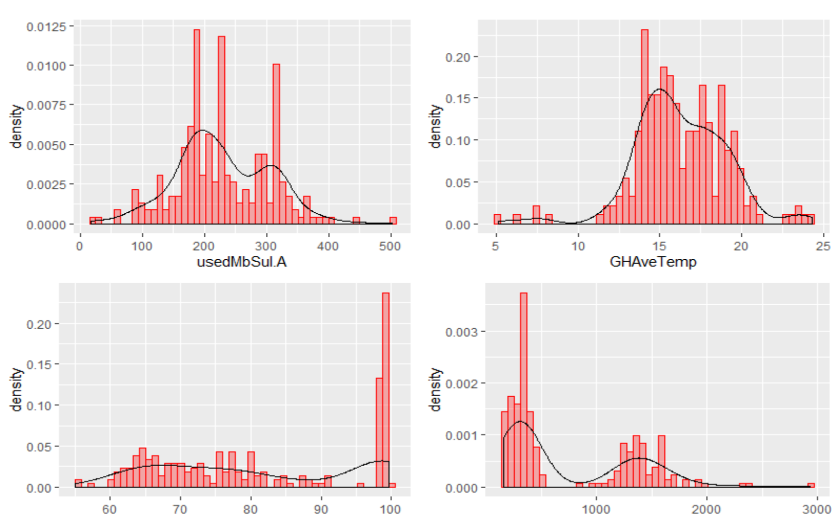

Figure 1. Real data of strawberry nutrition water quantity, temperature, humanity and CO_2 in Bed A of the strawberry greenhouse

Fig. 2 also shows the real data, which consist of strawberry nutrition water quantity, temperature, humanity and CO_2 sensed from the bed B of the strawberry greenhouse for 43 days. The scale of nutrition water quality is 94.47 to 328.05 litter used every day for 43 days. The total water is split into three varieties of the process. First, consider that monitoring in every day is a frequency of water quantity per day, this 0 value is the holiday. The production growth scale is 0.39 to 22.31 kg per day, while the nutrition water quality and production growth per cell in Bed B also proceeded for 43 days.

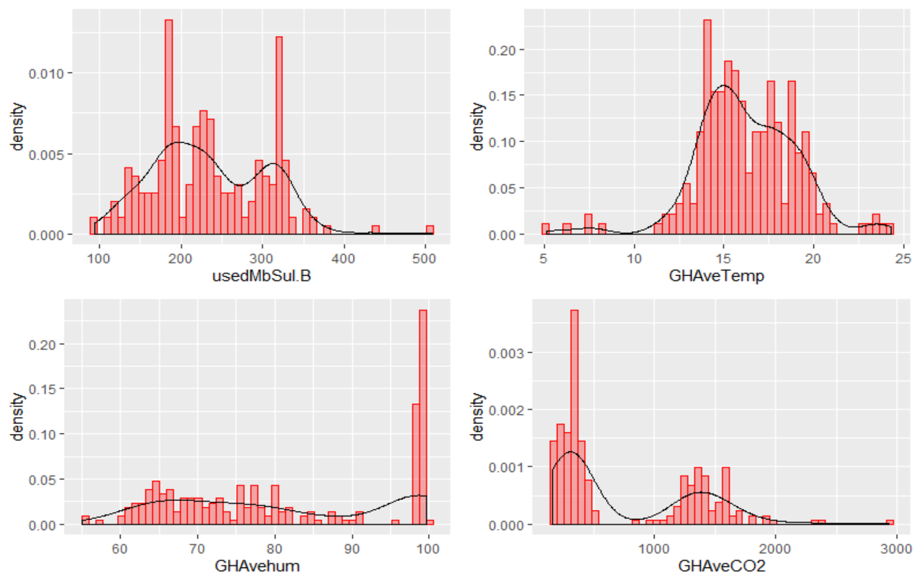

Figure 2. Real data of strawberry nutrition water quantity, temperature, humanity and CO_2 in Bed B of the strawberry greenhouse

4. Result and discussions

All 233 valid samples have been divided into training set and testing set. The partition rule of data is that the training set and the test set share similar statistical characteristics. Among them, 186 training sets are used in modeling while 46 test sets are used in examining the prediction performance of the model. The training sets account for 75% of all samples and cross-validation data is included in the training set. Three repetitive training and testing have been performed to obtain the average. The results are shown in table 1. The performance of the model has been examined from both aspects of the linear regression fitting and prediction accuracy of the model. Smaller the root means square error, which implies that the performance of the model is better and otherwise the model performs worse. The adopted evaluation indicators are: modeling root mean square (RMS) error, coefficient of determination coefficient R2 c; prediction root mean square error (RMSE), coefficient of determination R2 p, the ratio of the difference to the standard prediction error.

Fig.3 shows the result of strawberry water quality bed A and bed B. The correlation between compared with an equal relationship with everyone and nutrition water is positive water quantity per day. Thus, if nutrition water quantity per day increases then the water quantity per day cell fruit, will increase and ultimately strawberry growth will be increased. The water quantity per day bed plant in strawberry exhibited a positive significant association with the length of water quantity per day bed plant equal realization of water quantity per day cell fruit and growth. The positive significant association was recorded for the length of water quantity per day in strawberry with the nutrition of water cell fruit and growth per plant. Used two bed nutrition water in strawberry showed positive significant association with the growth of per plant. It is a parametric test, and assumes that the data are linearly related and that the residuals are normally distributed Section A 0.08779 and

Section B 0.07786683.

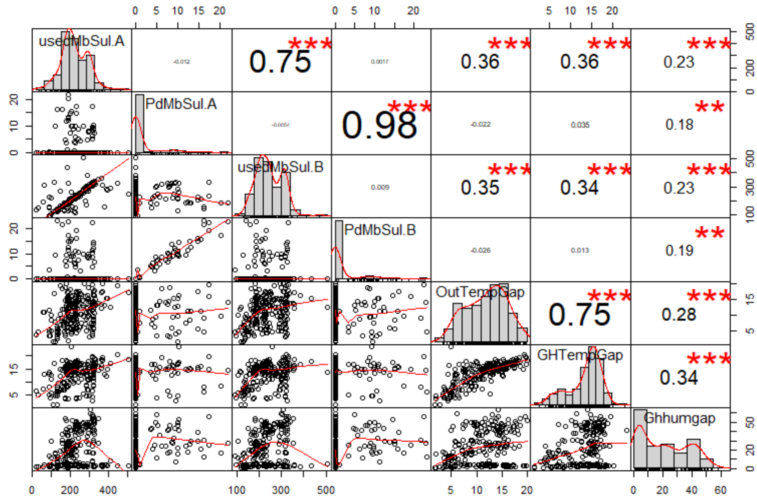

Figure 3. Correlation between Bed A and Bed B expressed temp, humanity, CO_2

It can be seen from table 1 that the linear regression model has better effects than bed B and bed A in the regression prediction of the growth rate of strawberries in the group of the same proportion. Regarding regression fitting degree and accuracy of the model samples, the linear model performs best with the largest coefficient of determination R2 c of 0.111, the smallest root mean square error RMSE of 56.07; performs worse in linear models with the smallest coefficient of determination and largest root mean square error; the best Section's effect after pre-processing the same data with relevant indicators superior to that of the traditional models.

Table 1. Performance indices R-squared, RMSE, P-value for LR model

Linear Regression Model		
	Bed A	**Bed B**
R-squared	0.126	0.11
RMSE	64.77	56.07
P-value	6.33	0.025

According to the principle of linear model evaluation and the comparison of the data in Table 1, when the training set ratio is 0.5, the regression fitting and prediction accuracy of the linear model is accurate. When the training data set ratio is 0.6 with the proportion of the training set increases, the prediction ability of various models has been improved, which the linear regression model is most significant. When the proportion of the training set is greater than or equal to 0.7, the linear regression model performs best to some extent. In general, the linear regression model can obtain more powerful information from increasing environmental information data, thus quickly enhancing model performance. It illustrates that, in the case of participation of multiple samples in the modelling, the linear regression model can significantly increase the regression fitting degree with much better performance than the traditional models. The model establishment for the transpiration rate of strawberry makes it happen that the growth transpiration rate can be predicted through the basic meteorological parameters in the greenhouse with high simulation accuracy and obtainable parameters. It can be a useful exploration of research on transpiration rate simulation in a short time scale, contributing to the more precise prediction based on Greenhouse temperature parameters.

Fig. 4 shows that bed A has F = 2.281 (p = 6.338) and bed B has F = 2.449 (p = 0.025), indicating that we should clearly reject the null hypothesis that the variables as like nutrition water quantity, average greenhouse temperature, average greenhouse humanity, and average greenhouse CO_2 collectively have no effect on price.

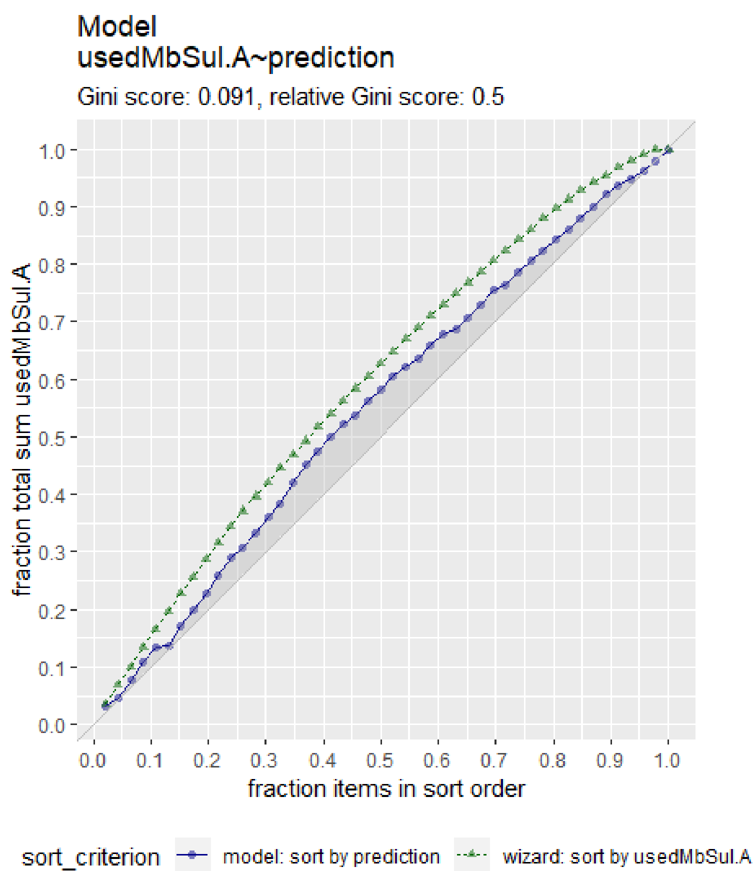

Figure 4. The deviation of the values prediction by LR model, from the actual value of strawberry Bed A

Fig. 5 also shows that the variable size is significant controlling for the variable water quantity (total) bed A (p = 6.33) and bed B (p = 0.025), as is water quantity controlling for the variable water quantity-bed (per day) bed A (p=0.1032) and section B (p=0.0210) and water quantity- cell bed A (p=0.0261) and bed B (p=0.025). In addition, the output also shows that bed A, R-squared = 0.126 and R-squared adjusted = 1, bed B R-squared = 0.0.111 and R-squared adjusted = 1. Careful scrutiny of the original data may reveal an error in data entry that can be corrected. The adjusted R square value is 1 which means the regression line is perfectly fitted in bed A with normal significance variable, the regression line is perfectly fitted in bed B with high significance variables. If they remain excluded from the final fitted model, they must be noted in the final result. Therefore, bed B gives the best prediction of strawberry growth production in this field.

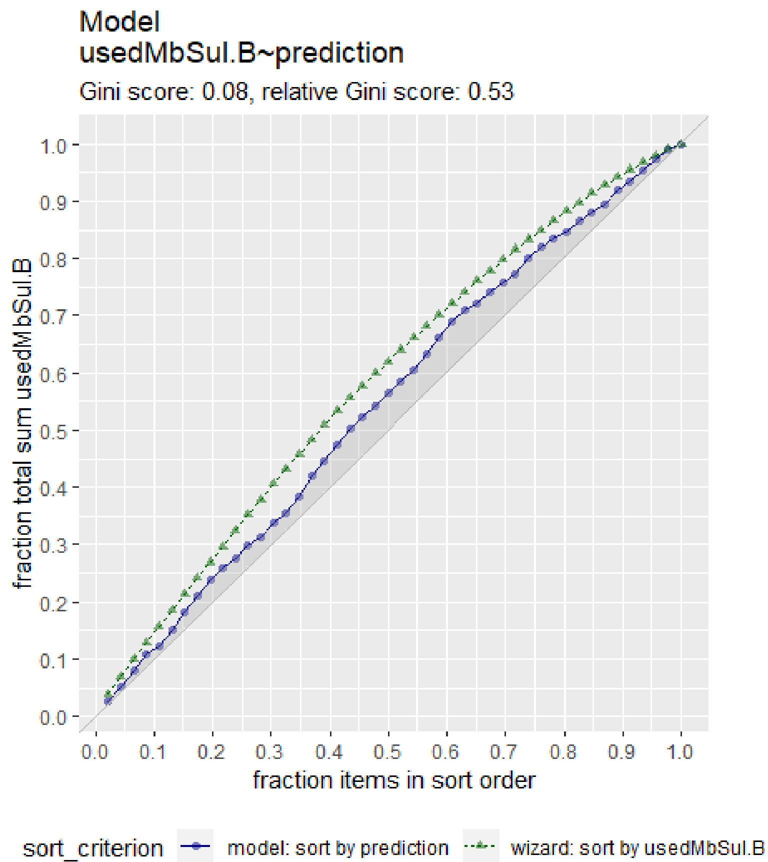

Figure 5. The deviation of the values prediction by LR model, from the actual value of strawberry Bed B

Furthermore, the prediction correlation result for each bed has habits compared and shown in table 1. The prediction coefficient of determination in the Linear Regression model continues to increase with testing and training sets, having the greatest accuracy in all models. The linear regression models are smaller with a case of sluggish growth. In general, the linear model can be obtained as a more powerful model from the increasing environmental information data, thus quickly enhancing model performance. It illustrates that, in the case of participation of multiple samples in the modeling, the linear model can significantly increase the regression fitting degree with much better performance than the traditional models. The linear model establishment for the nutrition water rate of strawberry makes it happen that the growth nutrition water rate can be predicted through the basic temperature parameters in the greenhouse with high simulation accuracy and obtainable parameters. It can be a useful exploration of research on nutrition water rate simulation in a short time scale, contributing to the more precise prediction based on temperature parameters.

Fig. 4 and 5 show for comparison between real data and they predicted the linear regression section bed A and section bed B. In section bed A, the blue dotted points show for real data and green dotted is sort by prediction the black line shows for regression line. The adjusted R square value is 0.126 which means the regression line is perfectly fitted in section bed A with normal significance variable. In section B, the blue dotted points show for real data and green dotted is sort by prediction the black line shows for the regression line. In section bed B also the adjusted R square value is 0.111. So, the regression line is perfectly fitted in section B with high significance variables. Thus, section A and section B are both the adjusted R square value is 0.1111. Therefore, the bed B gives the best prediction of strawberry growth production in this field

5. Conclusion

In this paper, we have used linear regression to find the better growth production of strawberry with optimum water quantity. Linear regression models are an effective approach for identifying the significant strawberry growth input intensity affecting water flow quantity and explaining the relationship between the greenhouse strawberry use intensity and the water quality, they appear to quantitatively estimate contribution of respective strawberry land-use intensity on the water quality because they only based on the existence of statistical significance in the analysis data. Our future research will focus on understanding water quantity, which increases the growth of the strawberry. We can interpret the possibility of this research progress in several ways. With a specific algorithm including the machine learning model, we plan to extend the reach of the output service. Considering that the machine learning model deals with reasoning, we can invent the linear regression method. The aim will be on understanding the amount of water required to increase the production of strawberries.

Acknowledgement

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01188605)" Rural Development Administration, Republic of Korea.

This research was supported by IPET (Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries) through Agri-Bio-industry Technology Development Program, funded by MAFRA (Ministry of Agriculture, Food and Rural Affairs) (No. 315001-5)

References

[1] Akın, M., and S. P. Eyduran. "Forecasting harvest area and production of strawberry using time series analyses." *Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi*, Vol. 34, No. 3, pp. 18-26, October 2017.
DOI: https://doi.org/10.13002/jafag4298

[2] Wrzesień, Mariusz, Waldemar Treder, Krzysztof Klamkowski, and Witold R. Rudnicki. "Prediction of the apple scab using machine learning and simple weather stations." *Computers and Electronics in Agriculture*, 161, pp. 252-259, June 2019.
DOI: https://doi.org/10.1016/j.compag.2018.09.026

[3] Shuaishuai, Li, Li Li, Meng Fanjia, Wang Haihua, Su Zhanzhan, and N. A. Sigrimis. "Prediction Model of Transpiration Rate of Strawberry in Closed Cultivation Based on DBN-LSSVM Algorithm." *IFAC-PapersOnLine*, Vol. 51, No. 17, pp. 460-465, January 2018.
DOI: https://doi.org/10.1016/j.ifacol.2018.08.171

[4] Liu, Changhong, Wei Liu, Xuzhong Lu, Fei Ma, Wei Chen, Jianbo Yang, and Lei Zheng. "Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit." *PloS one*, Vol. 9, No. 2, pp. 1-8 February 2014.
DOI: https://doi.org/10.1371/journal.pone.0087818

[5] ElMasry, Gamal, Ning Wang, Adel ElSayed, and Michael Ngadi. "Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry." *Journal of Food Engineering*, Vol. 81, No. 1 pp. 98-107, July 2007.
DOI: https://doi.org/10.1016/j.jfoodeng.2006.10.016

[6] Guo-Feng, P. A. N. "Research on spectral detection of nitrate in water quality based on K-S algorithm." *Chinese Journal of Spectroscopy Laboratory*, Vol. 5 , 2011.

[7] Jouquand, Céline, Craig Chandler, Anne Plotto, and Kevin Goodner. "A sensory and chemical analysis of fresh strawberries over harvest dates and seasons reveals factors that affect eating quality." *Journal of the American Society for Horticultural Science*, Vol. 133, No. 6, pp. 859-867, November 2008.

DOI: https://doi.org/10.21273/JASHS.133.6.859

[8] MacKenzie, S. J., and C. K. Chandler. "The late season decline in strawberry fruit soluble solid content observed in Florida is caused by rising temperatures." *Acta horticulturae*, Vol. 842, pp. 843-846, August 2009.
DOI: https://doi.org/10.17660/ActaHortic.2009.842.186

[9] Vicente, Esteban, Pablo Varela, Luis de Saldamando, and Gastón Ares. "Evaluation of the sensory characteristics of strawberry cultivars throughout the harvest season using projective mapping." *Journal of the Science of Food and Agriculture*, Vol. 94, No. 3, pp. 591-599, February 2014.
DOI: https://doi.org/10.1002/jsfa.6307

[10] Amid, Sama, and Tarahom Mesri Gundoshmian. "Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models." *Environmental Progress & Sustainable Energy*, Vo. 36, No. 2, pp. 577-585, March 2017.
DOI: https://doi.org/10.1002/ep.12448

[11] Jangmook Kang, and Sangwon Lee, “Algorithm Design to Judge Fake News based on Bigdata and Artificial Intelligence”, *IJIBC*, Vol. 11, No. 2, pp. 50-58, June 2019.
DOI: https://doi.org/10.7236/IJIBC.2019.11.2.50

[12] ByungJoo Kim, “Ensemble Methods Applied to Classification Problem”, *IJIBC*, Vol. 11, No. 1, pp. 47-53, March 2019.
DOI: https://doi.org/10.7236/IJIBC.2019.11.1.47