A study on the design of a melon fruit size prediction system using GDD and integrated solar radiation

1st Sangmin Lim
dept. Information and Communication Engineering
Sunchon National University
Suncheon si, Jeollanam-do
sangmin9642@gmail.com

2nd Meong Hun Lee
dept. Smart Agriculture Major
Sunchon National University
Suncheon si, Jeollanam-do
leemh777@scnu.ac.kr

3rd Hyun Yoe*
dept. Information and Communication Engineering
Sunchon National University
Suncheon si, Jeollanam-do
yhyun@scnu.ac.kr

*Corresponding author

***Abstract*—This paper focused on developing a system for predicting melon fruit size using the LSTM model. Melons are cultivated around the world in a variety of varieties with different rind shapes, colors, and fruit shapes, and net melons are mainly consumed in Korea. This study used an LSTM model to improve the accuracy of predicting melon fruit size, considering that melon cultivation area and production have an important impact on agricultural economy. The model was learned based on predictor variables such as number of days after modification, accumulated solar radiation, and effective integrated temperature, and the average MSE was 6.21. In addition, based on the research results, a monitoring system that can be used in melon farms was designed. This system is expected to contribute to improving agricultural productivity by providing data related to the growth process of melons, providing important information for harvest time and quality control. These studies highlight the positive impact of advances in agricultural technology and data-driven precision agriculture on the agricultural economy.**

Keywords—LSTM, Prediction, Melon, Growth, Data Analysis

I. INTRODUCTION

Melon (*Cucumis melo L.*) is an annual horticultural crop of the Cucurbitaceae family. Recently, various varieties of melon have been improved and are now cultivated all over the world. Melons have different rind shapes, fruit shapes, and colors depending on the variety, and net melons are mainly consumed in Korea[1]. As of 2021, Korea's melon cultivation area is 1,518 ha, production is approximately 41,000 tons per year, and production amount is approximately KRW 120 billion[2]. Melon has a wide consumer base in Korea due to its sweet taste and unique aroma, and is a promising crop for export as it is the second largest fruit and vegetable export unit price (USD 4.49/kg)[3].

Recently, the preference for good quality, safe and healthy foods is growing in Korea, and the same goes for fruits. In addition, it is increasing in a wide range of consumption forms and is rapidly changing into high value-added products such as various flesh and skin colors and seed-related functional melons[4]. Melon consumption continues to increase, and its quality and production have a significant impact on the agricultural economy. In particular, the fruit size of a melon is a critical factor that directly affects market value and consumer satisfaction. However, predicting melon fruit size is complicated by various environmental factors, and accurate prediction requires specialized knowledge and experience that cannot be made by ordinary farmers.

In this paper, we design a system that uses effective integrated temperature and integrated solar radiation and utilizes an LSTM model to more accurately predict melon fruit size and monitor the predicted results in melon farms. This system provides melon farmers with important information for harvest timing and quality control, and can improve agricultural productivity. The goal of this paper is to help melon farmers establish more accurate production plans. Additionally, this system is expected to contribute to improving the accuracy of melon quality control and yield prediction.

II. RELATED RESEARCH

A. Introduction of LSTM

In this paper, we use the LSTM model to predict melon fruit size. The LSTM model is a time series data-based recurrent neural network model that can process data by learning dependencies between data[5]. LSTM is a model that emerged after RNN, and the existing RNN has the problem of not being able to utilize past information well when the distance between the information to be used and the information to be predicted increases during the learning process. The reason for this is that learning proceeds by sequentially multiplying gradients in the backpropagation process, and as the layer gets deeper, there is a possibility that the value will converge to 0 as more gradients are multiplied. Therefore, the LSTM model was developed to improve this problem. LSTM can efficiently transmit information over a relatively long distance by using forget gates and input gates to decide whether to discard information or not and how much to use. Figure 1 shows the basic structure of the LSTM model[6].

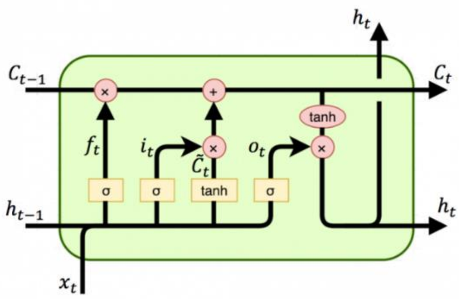

Figure 1. Basic structure of LSTM model

III. MAIN SUBJECT

A. System architecture design

The system to be designed in this paper is shown in Figure 2. Smart greenhouses measure environmental information and growth data. The collected data is stored in a database, goes through a preprocessing process, and is then trained as an LSTM model. The performance indicator of prediction accuracy was MSE (Mean Squared Error). MSE is one of the deep learning loss functions, and the smaller the result, the better the performance. The MSE formula is as Equation (1).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{Y_i} - Y_i)^2 \quad (1)$$

The predicted values and MSE values resulting from the learning results of the LSTM model are provided to users through a web page or smartphone. Based on the information provided, users can receive help in harvesting time or controlling the smart greenhouse environment.

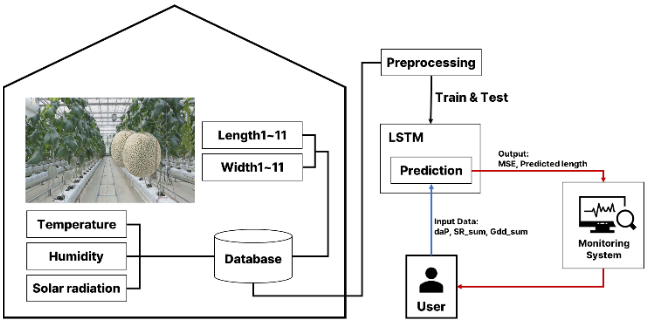

Figure 2. Fruit size prediction system structure diagram

B. Data types and collection

In this paper, environmental information data and growth information data are stored in a database. Types of environmental information data include greenhouse internal temperature, internal humidity, external temperature, and solar radiation information stored on a minute-by-minute basis. Growth information data measures the length and width of a total of 11 fruits from fertilization to harvest. Measurements are made directly by humans at intervals of approximately one week and stored in the database. Growth data was used for melons revised on April 24, 2023 and melons revised on August 12, 2023. TABLE I shows the sensors used to collect environmental information data and the items of the collected data.

TABLE I. SENSORS USED AND DATA ITEMS COLLECTED

	Sensors used for collection and data items collected		
	Data	***Sensor***	***Unit***
in_temp	Internal temperature	NTC10K	˚C
out_temp	Outside temperature	NTC10K	˚C
in_humi	Internal Humidity	HCPV-220	%
SR_d	Integrated solar radiation	WSS-202	MJ/m^2
daP	Number of days since plant	-	-
length1~11	length of fruit	-	mm
width1~11	width of fruit	-	mm

C. Data preprocessing

Since the collected environmental information data is in minutes, daily averages for temperature and humidity were calculated to match them to daily values. In addition, the effective integrated temperature (GDD) and integrated solar radiation were calculated based on the collected environmental information data. The formula for calculating the effective integrated temperature is Equation (2).

$$GDD = \sum T_{avg} - T_{base}, \ T_{base} = 15 \quad (2)$$

The accumulated solar radiation was calculated by accumulating the accumulated solar radiation from the 1st day after modification to 23:59. After calculating daily average temperature and humidity data, integrated solar radiation, and effective integrated temperature, data normalization was performed to reduce scale differences between characteristics. Min-Max Normalization was used as a normalization method.

D. LSTM model training

The data used in this paper are melon growth information data and environmental information data for a total of two growing seasons. Since the next cropping period does not start immediately after the first cropping period, time series characteristics were created by distinguishing between the first cropping period and the second cropping period. Figure 3 is a graph dividing the first and second periods by counting data points by date. As can be seen in the graph, the first data collection period is from May 4, 2023 to June 19, 2023, and the second data collection period is from August 21, 2023 to October 5, 2023.

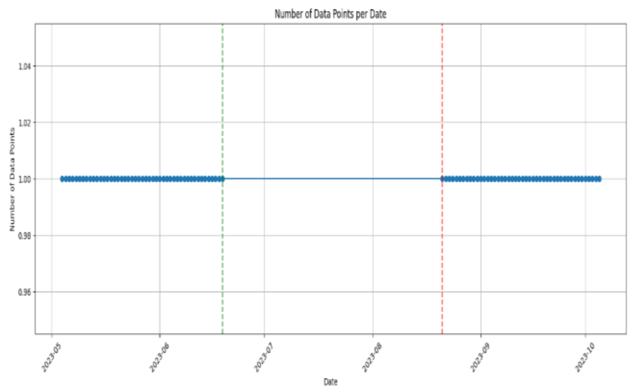

Figure 3. Division graph of the first and second operations

In addition, length1 to 11 (exaggeration) of a total of 11 fruits were set as target variables, and correlations were analyzed to determine predictor variables, and characteristics with a correlation coefficient of 0.6 or more were set as predictor variables. Figure 4 is a heatmap graph used for correlation analysis. In the correlation analysis, the exaggerations of the 11 fruits were not used, but only one representative exaggeration was used for analysis. As a result of the correlation analysis, daP (days after correction), SR_sum (integrated solar radiation), and gdd_sum (effective integrated temperature) were used as predictor variables.

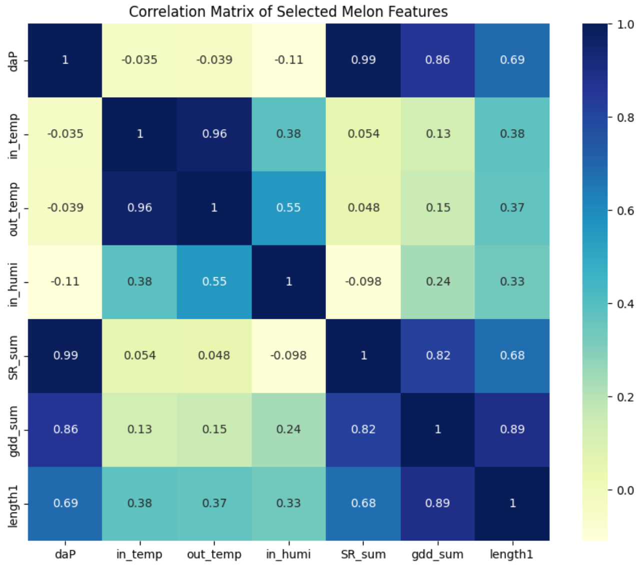

Figure 4. Correlation heatmap analysis to set predictor variables.

Because the amount of data was limited, the past 3 days of data was used to predict the future. In addition, since there are two cropping seasons, the data from each cropping season was processed to generate time series characteristics to predict the next day's melon length based on 3 days of data, and then integrated to create one dataset. Training data and test data were divided 8:2. relu was used as the LSTM activation function, and the number of neurons in the output layer was set to 11, which is the number of target variables. The epoch was learned at 200 epochs.

E. *LSTM model training results*

As a result of learning the LSTM model, the MSE of the first operation was 7.31 and the MSE of the second operation was 5.11. Figure 5 is a graph of the prediction results of the first operation, and Figure 6 is a graph of the prediction results of the second operation.

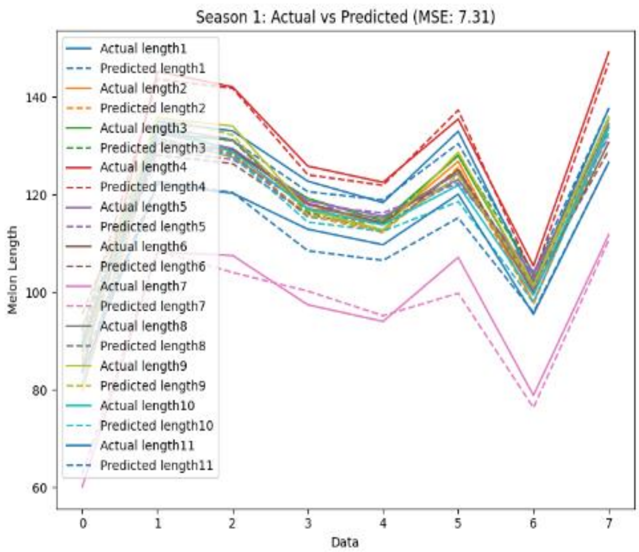

Figure 5. Predicted results of the first planting period

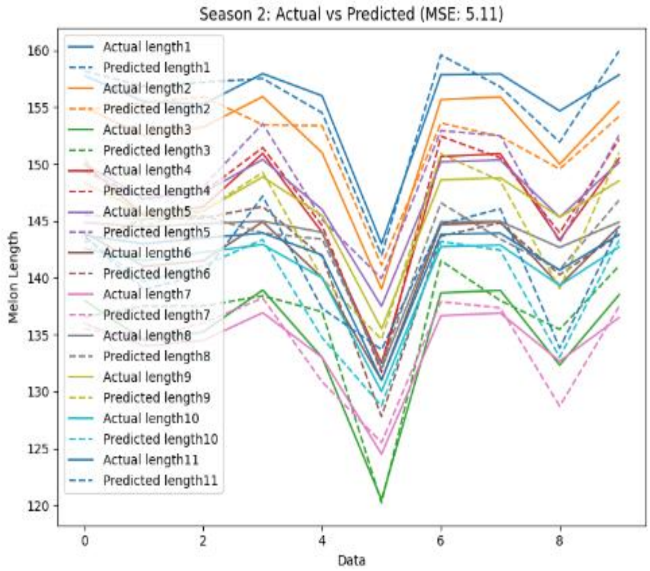

Figure 6. Predicted results of the second cropping period

F. *Monitoring system design*

A monitoring system was designed using the learned model. The monitoring system was created as a web page, designed based on ubuntu-22.04, and written based on the open source grafana. The user can obtain the predicted exaggerated value through the LSTM model by entering daP, SR_sum, and gdd_sum as input values. In addition, the collected data and predicted values can be displayed in a graph to view the melon's growth process in a graph. Through the monitoring system, users can check the environmental information necessary for melon growth and can control the environment accordingly. In addition, it is believed that the harvest time can be determined based on future prediction results and efficient production will be possible.

IV. CONCLUSION

In this study, the LSTM model was used to accurately predict melon fruit size. LSTM is a deep learning model that can learn the long-term dependency of time series data and is suitable for predicting changes in data over time. This model, based on melon growth data and environmental data, identifies important variables related to the melon growth process and predicts fruit size through these. In this study, the number of days after correction, accumulated solar radiation, and effective integrated temperature were used as predictors, and the average MSE was 6.21. Additionally, it can make an important contribution to improving the cultivation efficiency and productivity of melon farmers by designing a system that notifies the prediction exaggeration when the user inputs the predictor variables. Farmers can use this system to more precisely manage harvest time and quality, and can expect increased production and improved quality in the long term. Additionally, this study shows the positive impact of advances in agricultural technology and data-driven precision agriculture on the agricultural economy.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2024-2020-0-01489) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

REFERENCES

[1] M.Y. Lim, S.H. Choi, G.L. Choi, S.H. Kim, and H.J. Jeong 2020, Growth and quality of muskmelon (Cucumis melo L.) as affected by fruiting node order, pinching node order and harvest time in hydroponics using coir substrate. Protected Hort Plant Fac 29:406-413. (in Korean) doi:10.12791/KSBEC. 2020.29.4.406

[2] Ministry of Agriculture, Food and Rural Affairs (MAFRA) 2022b, Present status of greenhouse and vegetable production in 2022. MAFRA, Sejong, Korea, p 116.

[3] Ministry of Agriculture, Food and Rural Affairs (MAFRA) 2022a, Agricultural business registration information statistics service. Available via https://uni.agrix.go.kr/

[4] J.Y. Baek, J.I. Park, Y.H. Kim and I.H. Lee, "Ministry of Agriculture, Food and Rural Affairs final report on breeding high-quality melon varieties customized for export to Japan, etc.," National Library of Korea, 2018.

[5] M. Zhu, X. Guan, Z. Li, L. He, Z. Wang, and K. Cai, "sEMG-based lower limb motion prediction using CNN-LSTM with improved PCA optimization algorithm," Journal of Bionic Engineering, vol. 20, no. 2, pp. 612-627, 2023

[6] LSTM, https://medium.com/dovvie/deep-learning-long-short-term-memory-model-lstm-d4ee2f005973