Estimation of Greenhouse Tomato Transpiration through Mathematical and Deep Neural Network Models Learned from Lysimeter Data

Meanne P. Andes[1], Mi-young Roh[2], Mi Young Lim[3], Gyeong-Lee Choi[3], Jung Su Jung[4], and Dongpil Kim[3]*

[1]*Engineer II, Bureau of Plant Industry-Los Baños National Crop Research Development and Production Support Center 4030, Philippines*
[2]*Senior Researcher, Protected Horticulture Research Institute, NIHHS, RDA, Haman 52054, Korea*
[3]*Researcher, Protected Horticulture Research Institute, NIHHS, RDA, Haman 52054, Korea*
[4]*Rural Extension Instructor, Gyeongsangbuk-do Provincial Agricultural Research and Extension Services, Daegu 41404, Korea*

Abstract. Since transpiration plays a key role in optimal irrigation management, knowledge of the irrigation demand of crops like tomatoes, which are highly susceptible to water stress, is necessary. One way to determine irrigation demand is to measure transpiration, which is affected by environmental factor or growth stage. This study aimed to estimate the transpiration amount of tomatoes and find a suitable model using mathematical and deep learning models using minute-by-minute data. Pearson correlation revealed that observed environmental variables significantly correlate with crop transpiration. Inside air temperature and outside radiation positively correlated with transpiration, while humidity showed a negative correlation. Multiple Linear Regression (MLR), Polynomial Regression model, Artificial Neural Network (ANN), Long short-term Memory (LSTM), and Gated Recurrent Unit (GRU) models were built and compared their accuracies. All models showed potential in estimating transpiration with R^2 values ranging from 0.770 to 0.948 and RMSE of 0.495 mm/min to 1.038 mm/min in the test dataset. Deep learning models outperformed the mathematical models; the GRU demonstrated the best performance in the test data with 0.948 R^2 and 0.495 mm/min RMSE. The LSTM and ANN closely followed with R^2 values of 0.946 and 0.944, respectively, and RMSE of 0.504 m/min and 0.511, respectively. The GRU model exhibited superior performance in short-term forecasts while LSTM for long-term but requires verification using a large dataset. Compared to the FAO56 Penman-Monteith (PM) equation, PM has a lower RMSE of 0.598 mm/min than MLR and Polynomial models degrees 2 and 3 but performed least among all models in capturing variability in transpiration. Therefore, this study recommended GRU and LSTM models for short-term estimation of tomato transpiration in greenhouses.

Additional key words: tomato transpiration model, regression model, Artificial Neural Network, LSTM, GRU

Introduction

Transpiration is a fundamental process involving losing water as water vapor, starting from the tiny organ, stomata in their leaves. Nearly all water taken up by plants is lost by transpiration, and only a small fraction is utilized. This process occurs simultaneously with evaporation, called evapotranspiration (ET). Since they occur simultaneously, it is difficult to distinguish between the two processes. However, during the early stages of crop growth, all water loss is attributed to evaporation, while during full crop cover, more than 90% is due to transpiration (Allen et al.,

1998). The water movement from transpiration plays a vital role in maintaining the water balance of plants (Hazlett, 2022). This equilibrium is important to prevent dehydration and water stress in the short term and to support the growth and production of fruits and flowers in the long-term perspective. To attain equilibrium, water uptake from the root zone must equal the evaporation rate (Geelen et al., 2020).

Sufficient water can be provided for the longer term if irrigation needs are aligned with the evaporation energy received by the plant (Geelen et al., 2020). Among the points that contribute to plant evaporation are temperature, humidity, air movement, and light intensity. Increasing or decreasing the level of these energy inputs can either increase or decrease the transpiration rate (PASSeL, 2023). Irrigation is essential to producing most vegetables to attain

---

good and high-quality yields. Vegetables such as cucumber, tomato, lettuce, zucchini, and celery have a very high percentage of water content in the cells, thus extremely vulnerable to water stress and drought conditions (Yildirim and Ekinci, 2022). However, over-irrigation can inhibit germination and root development and decrease vegetable quality and post-harvest life (Yildirim and Ekinci, 2022). Regardless of the crop yield, the importance of appropriate irrigation technology is increasing regarding resource conservation in hydroponics.

One way of obtaining plant water demand is by measuring water lost by transpiration (Sanchez et al., 2012). In crop cultivation under a greenhouse with water use efficiency (WUE) reported to be three to 10 times higher than under open field conditions, evapotranspiration knowledge may help improve the plant environment and WUE (Katsoulas and Stanghellini, 2019). The lower evaporative demand inside the greenhouse than the open field reduces water requirement and consequently increases water use efficiency (Gallardo et al., 2013). Furthermore, being at the forefront of "precision agriculture", greenhouses increasingly need precision irrigation and climate management. Hence, knowledge of crop transpiration at relatively short intervals (hours and minutes) is necessary (Katsoulas and Stanghellini, 2019). Measuring transpiration on a time scale using weighing lysimeters or sap flow measurement takes more time and costs, so crop transpiration models are commonly adopted (Katsoulas and Stanghellini, 2019). The common models for predicting evapotranspiration and transpiration are Penman-Monteith (PM) (Allen et al., 1998), Shuttleworth-Wallace (SW) (Shuttleworth and Wallace, 1985), and Priestly-Taylor (PT) (Priestly and Taylor, 1972) models (Shao et al., 2022). PM is the most recommended worldwide standard method because it integrates energy, aerodynamic, and atmospheric parameters (Chia et al., 2022). Other mathematical models that include regression models, such as linear, exponential, logarithmic, polynomial, and power, have been used to estimate the evaporation and transpiration of crops such as Maize (Saedi, 2022). Multiple linear regression (MLR) (Tu et al., 2019; Bera et al., 2021; Li et al., 2023) was also used to predict transpiration in canopies. However, applications of mathematical models are still limited because their parameterization is very complex and needs a large number of observation data (Fan et al., 2021)

and thus impractical in regions where data collection facilities are incomplete (Chia et al., 2022).

Recently, machine learning models have been successfully used to estimate evapotranspiration withlimited meteorological data (Ferreira and da Cunha, 2020). These models can capture complex relationships between input and output data, thus making them powerful tools in evapotranspiration modeling (Ferreira and da Cunha, 2020). Moreover, the machine learning techniques can capture hydrological time series such as evapotranspiration by utilizing solely a series of predictors without any knowledge of their physical processes (Mehdizadeh et al., 2021; Mohammadi and Mehdizadeh, 2020; Mohammadi et al., 2021; Elbeltagi et al., 2021). Several machine learning models to estimate the transpiration of different crops were assessed, such as artificial neural network (ANN) (Ferreira and da Cunha, 2020; Yong et al., 2023; Tunali et al., 2023), convolutional neural network (CNN) (Ferreira and da Cunha, 2020; Li et al., 2023), long short-term memory (LSTM) (Chen et al., 2020; Chia et al., 2022; Li et al., 2023), gate recurrent unit (GRU) (Chia et al., 2022; Li et al., 2023). The studies above showed the promising performance of machine learning models in estimating transpiration.

Despite several studies conducted to predict transpiration using different methods, few studies have been conducted comparing the prediction performance of mathematical (MLR and polynomial) models and deep learning (ANN, LSTM, and GRU) models using data on smaller time scales (every minute). To reduce overestimated irrigation amount, instantaneous transpiration with shorter intervals is more favorable than daily accumulated, conventionally applied for irrigation (Shin et al., 2014). Hence, this study aims to estimate tomato transpiration through mathematical and deep learning models using every minute data and to identify the suitable model.

Materials and Methods

1. Data Gathering, Preprocessing, and Analysis

Tomatoes (*Solanum lycopersicum* L.) were grown inside the Venlo-type greenhouse at Protected Horticulture Research Institute (PHRI), Haman, Republic of Korea. Seeds were sown on October 11, 2022, and transplanted in coconut coir

Meanne P. Andes · Mi-young Roh · Mi Young Lim · Gyeong-Lee Choi · Jung Su Jung · Dongpil Kim

substrate on November 17, 2022. Every minute data of parameters which include weight of the plant and substrate (kg), slab temperature (°C), volume of irrigation (mL) and drain (mL), inside air temperature (°C), humidity (%), electrical conductivity (mS/cm), pH, outside air temperature (°C), and solar radiation (W/m^2) were gathered. Every minute, a load cell was used to get the crop weight data (Incrocci et al., 2020). The transpiration rate of tomatoes was obtained based on the weight change of plants in each unit of time (Shin et al., 2014). Weight changes due to crop management, such as pruning, harvesting, and evaporation at the surface of the substrate, were not included in transpiration (Jo and Shin, 2021; Shin and Son, 2015). Data considered in this study were those obtained from January 2, 2023, to May 2, 2023. These were imported in Python 3.10.9 for preprocessing, analysis, and model building. Missing values of outside radiation were interpolated with a linear method in the Python library Pandas 2.1.1. Moving average calculations were performed to smooth out fluctuations in data. A 30 window (number of data points) was applied to transpiration, and a 50 window for outside radiation. The strength and direction of a relationship between transpiration and dependent variables (slab temperature, volume of water used, inside air temperature, humidity, EC, pH, outside air temperature, and solar radiation) were obtained using the Pearson correlation coefficient. A two-tailed correlation analysis at a 1% significance level was done using SciPy. Stats package from SciPy library in Python. Scatter plots of transpiration versus dependent variables were plotted to visualize the relationship. The relationship between transpiration and independent variables was also observed by plotting the line plot of variables over time. For model building, the data were split into training (70%), validation (15%) and testing (15%) datasets. Inputs were normalized using MinMaxScaler to improve training efficiency.

2. Model Building

Seven models were developed to estimate tomato transpiration: multiple linear regression model (MLR), polynomial regression models degrees 2, 3, and 4, Artificial Neural Network (ANN), long short-term model (LSTM), and Gated Recurrent Unit (GRU) model. To get the best architecture, trial and error were performed in building ANN,

LSTM, and GRU models. All models were built using applicable features, classes, and libraries in Python within the Jupyter Notebook environment.

2.1 Multiple Linear Regression (MLR) Model

MLR is an extension of simple linear regression, which estimates the relationship between a response variable y and an independent variable x. However, the MLR model is extended to include more than independent variables (x_1, x_2, ...x_p), producing a multivariate model (Tranmer et al., 2020). In this model, we assumed that the dependent variable is directly related to a linear combination of the independent variables. The equation (Eq.1) for MLR has the same form as that for simple linear regression but has more terms:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \qquad (1)$$

where Y_i - dependent variable
β_o - intercept or constant
β_1, β_1, ... β_k - slope of regression surface
x_{i1}, x_{i2}, ...x_{ik} - independent variables
e_i - error term

2.2 Polynomial Regression Model

Polynomial regression is a special case of multiple regression, in which the relationship between the independent and dependent variables is modeled in the nth-degree polynomial (Ostertagová, 2012). This model is useful when the relationship between variables is curvilinear (Ostertagová, 2012). This can be expressed in the following equation (Eq. 2):

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 \cdots + \beta_k x_i^k + e_i \qquad (2)$$

where Y_i - dependent variable
β_o - intercept or constant
$\beta_1, \beta_2, \ldots \beta_k$ - slope of regression surface
x_i - independent variables
k - degree of polynomial
e_i - error term

2.3 Artificial Neural Networks (ANN)

ANNs are biologically inspired computational networks (Park and Lek, 2016). These are commonly presented as

interconnected "neurons" systems that can compute input values by feeding information through the network (Dahikar and Rode, 2014). Typically, a minimum of three layers consisting of input, hidden, and output layers is required in developing ANN, but hidden layers can be extended depending on specific problems (Bejo et al., 2014). In this study, all deep learning architectures were implemented and trained using the PyTorch 2.0 framework. The feed-forward neural network was implemented where information flows from the input layer through hidden layers to the output layer. The ANN has seven input layers and two hidden layers, with 28 nodes in the first hidden layer and 14 nodes in the second. The output layer has only one node, which is the transpiration rate. Rectified linear unit (ReLU) was the activation function used from the input layer to the first and second layers, while linear activation function from the second layer to the output layer. To reduce losses and provide the most accurate results, an adaptive moment estimation (Adam) optimizer was used in the training (Chauhan, 2020). Early stopping at a patience of 20 was applied to prevent overfitting and improve generalization (Vanja et al., 2021). The training was performed for 100 epochs with a batch size of 64.

2.4 Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)

LSTM architecture is a special recurrent neural network (RNN) with an appropriate gradient-based learning algorithm to overcome error backflow problems (Tian et al., 2020). It has chain-like modules wherein each repeating module contains a memory block designed to store information over long periods (Zhang et al., 2018). The memory block comprises four parts: the cell state or CEC (Constant Error Carousel) and three special multiplicative units called gates. The input, forget, and output gates in each memory block can control the flow of information inside the memory block (Zhang et al., 2018). The forget gate decides which values of the previous cell state should be discarded and which should be kept. Then, the input gate selects values from the last hidden state and the current input to update by passing them through the sigmoid activation function. The cell state candidate regulates the flow of information in the network by using an activation function on the previous hidden state

and current input. The candidate calculated in the cell state is then added to the previous cell state. Lastly, the output gate calculates the current hidden state by passing the previous hidden state and the current input through the sigmoid activation function to select which new information should be considered. The current cell state value is then passed through the tanh function, and the products of those functions are multiplied. During backpropagation training, the gradient flow is relatively undisturbed because only a few linear operations are performed on the cell state, hence limiting the occurrence of the vanishing gradient problem (Zarzycki and Lawrynczuk, 2021).

In this study, before initializing the LSTM model, the input sequence length was set to 60 time steps. The LSTM structure has two LSTM layers, with the first layer containing 28 memory cells and the second layer containing 14. The output layer has a single neuron of transpiration. The model was trained at a learning rate 0.001 and optimized using Adam optimizer. Early stopping at Patience 20 was also applied. The training was performed for 100 epochs with a batch size of 64.

GRU is an RNN gating mechanism like LSTM but with only two gates, namely, reset gate and update gate, thus higher computational efficiency and faster convergence. GRU structure is also more concise and includes fewer parameters than LSTM, which minimizes overfitting and improves training efficiency (Li et al., 2022). The reset gate selects which information to discard from the previous hidden state and input values, while the update gate selects which information from the previous hidden state should be kept and passed along to the next steps. The candidate state gate calculates the candidate for a future hidden state by multiplying the previous state with the reset gate's output. Then, the new data from the input is added to the remaining information. Finally, the tanh function is applied to the data to regulate the information flow. The length of the input sequence was set to 60 time steps before initializing the GRU model. The GRU structure has two GRU layers, with the first layer containing 28 neurons and the second layer containing 14 neurons, while the output layer has a single neuron that yields the output, which is transpiration. The model was trained at a learning rate 0.001 and optimized using Adam optimizer. Early stopping at Patience 20 was

Meanne P. Andes · Mi-young Roh · Mi Young Lim · Gyeong-Lee Choi · Jung Su Jung · Dongpil Kim

also applied. The training was performed for 100 epochs with a batch size of 64.

2.5 Calculation of transpiration using Penman–Monteith (PM) Equation

The FAO56 PM equation has been generally recommended as the standard method for estimating evapotranspiration (ET_o) for most crops and cropping systems (Incrocci et al., 2020; Allen et al., 1998). The FAO-styled "reduced form" of PM equation results that has also been adopted by ASCE-EWRI in 200 was considered in the calculation of evapotranspiration (Allen et al., 2006) (Eq. 5)

$$ET_o = \frac{0.408\triangle(R_n + G) + \gamma\frac{C_n}{T+273}u_2(e_s - e_a)}{\triangle + \gamma(1 + C_d u_2)} \qquad (5)$$

where ET_o - evapotranspiration (mm/hr)

$\triangle$ - slope vapor pressure curve (kpa/°C)

R_n - net radiation at crop surface (MJ/m²/hr)

G - soil heat flux density (MJ/m²/hr)

T - mean hourly air temperature at 2 m height (°C)

u_2 - wind speed at 2m height,

$u_2 = 208/ra$; ra = 295 s/m; ra - aerodynamic resistance (Fernandez et al., 2011)

e_s - saturation vapor pressure (kPa)

e_a - actual vapor pressure (kPa)

C_n - crop coefficient, Cn=900 (24-h), Cn = 37 (hourly time steps)

C_d - daylight coefficient, Cd= 0.34 (Allen et al., 1998)

To get crop transpiration, Eq. 6 was used:

$$ET_c = ET_o * K_c \qquad (6)$$

where k_c is the crop coefficient, the crop coefficient values for tomatoes ($K_{cmax} = 1.4$, $K_{cend} = 0.85$) grown in a Mediterranean greenhouse obtained by Magan et al. (2008) were used. Transpiration calculated using the PM equation was compared to the measured transpiration. The R^2 and RMSE between measured and calculated values were obtained and compared to the R^2 and RMSE of the models.

2.6 Model Evaluation

Coefficient of Determination (R^2) and Root Mean Square Errors (RMSE) were used to evaluate the models' performance. The evaluation was conducted using NumPy and Scikit-Learn (Sklearn) library in Python. The values of R^2 and RMSE of the models developed were compared to determine the best-fitted model to estimate tomato transpiration.

Results and Discussion

1. Correlation between the environmental variables

The Pearson correlation coefficient of the variables showed a significant correlation of all measured independent variables (humidity, inside air temperature, outside radiation, EC, pH, slab temp, outside temperature, irrigation, drain, month, and hour) with the transpiration except for day and minute (Table 1). Outside radiation and inside air temperature showed the highest significant positive correlation with transpiration, with correlation coefficients of 0.793 and 0.725, respectively. This indicates that increased outside radiation and inside air temperature can significantly increase transpiration (De Wit, 1958; Jolliet and Bailey, 1992; Zhu et al., 2022). In terms of humidity, a negative correlation shows that an increase in humidity leads to a decrease in transpiration rate (Zhu et al., 2022). The result of correlation analysis was

**Table 1.** Correlation coefficient of independent variables.

Variables	Correlation Coefficient
Humidity	− 0.297*
Inside air temperature	0.725*
Outside radiation	0.842*
EC	0.113*
pH	0.168*
Slab temperature	0.277*
Outside air temperature	0.165*
Irrigation	0.023*
Drain	0.023*
Month	− 0.188*
Day	− 0.004
Hour	− 0.074*
Minute	0.003

*Significant at p <0.01 level (2-tailed).

used as a basis for selecting input features for building estimation models to ensure their relevance in predicting the target variables. However, not all variables with significant correlation were chosen for model building. This study considered the first three variables with the highest correlation (outside radiation, greenhouse air temperature, and humidity).

Furthermore, air temperature, relative humidity, and radiation are environmental variables that greatly influence crop transpiration (Jo and Shin, 2021). Fig. 1 shows the scatter plot depicting the relationship trend between independent variables and transpiration. In terms of month and

hour, despite a significant correlation, the direction of the relationship does not perfectly reflect the actual scenario. Hence, the decision to include these variables, including the day and minute in model building, was based on the observed seasonality of transpiration over time . This is based on the assumption made in forecasting that there is an underlying pattern which described the event and conditions and that it repeats in the future. Identification of patterns and choice of model, particularly in time series, is critical to facilitate forecasting (Nwogu et al., 2016).
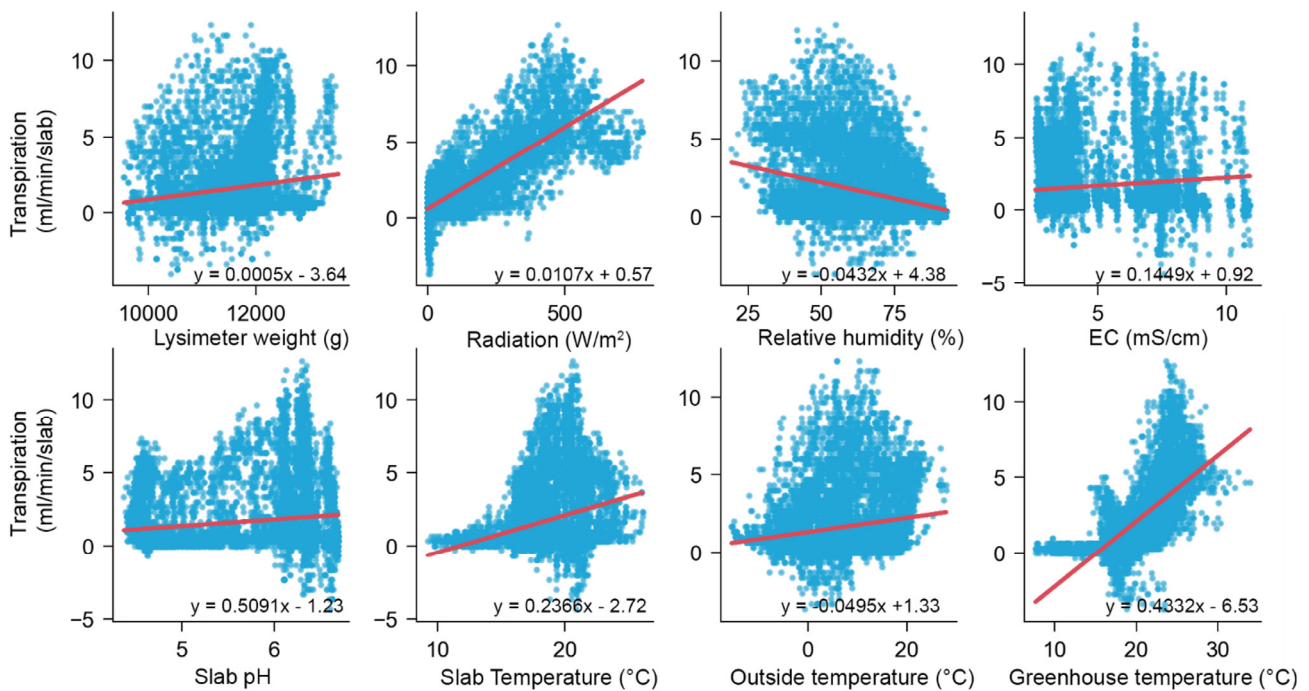
Fig. 1. Correlation between the collected variables and transpiration within the greenhouse. Individual data points are represented by blue dots, while the regression line is depicted in red.

Table 2. Model-wise RMSE and R^2 values across training, validation, and testing phases.

Model	Training		Validation		Testing	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
MLR	0.765	1.061	0.767	1.046	0.770	1.038
Polynomial Degree 2	0.866	0.800	0.867	0.791	0.869	0.783
Polynomial Degree 3	0.905	0.673	0.905	0.668	0.906	0.663
Polynomial Degree 4	0.931	0.575	0.931	0.568	0.932	0.565
ANN	0.946	0.510	0.945	0.509	0.944	0.511
LSTM	0.948	0.497	0.946	0.502	0.946	0.504
GRU	0.952	0.480	0.948	0.496	0.948	0.495

Meanne P. Andes · Mi-young Roh · Mi Young Lim · Gyeong-Lee Choi · Jung Su Jung · Dongpil Kim

2. Comparison of Model Performance

Models built along with the performance evaluation results are shown in Table 2. All models showed potential in estimating transpiration rate using data on radiation, temperature, humidity, and time with R^2 values ranging from 0.770 to 0.948 and RMSE of 0.495 mm/min to 1.038 mm/min in the testing. During training and validation, the R^2 and RMSE of the models did not go far from each other, with an R^2 range of $0.001 - 0.004$ and $0.001 - 016$ for RMSE; hence, overfitting was successfully managed. Early stopping, specifically in the building ANN, LSTM, and GRU models, helped prevent overfitting and reduced its effects on the model (Ying, 2019). For mathematical models, results showed that Polynomial models degrees 2, 3, and 4 have better estimation performance in the testing compared to the MLR model (R^2 = 0.77, RMSE= 1.038mm/min), hence showing that the relationship between the independent variables and transpiration is more curvilinear than linear. Among the regression models, polynomial degree 4 (P4) model showed best performance (R^2 = 0.93, RMSE = 0.565 mm/min). The polynomial model is useful when the relationship between variables is curvilinear (Ostertagová, 2012).

Meanwhile, the performance of deep learning models such as ANN, LSTM, and GRU was better than that of mathematical models. Unlike deep learning models, which do not require knowledge of internal factors and can be constructed with limited data, mathematical models need observation data (Fan et al., 2021). The addition of more relevant input variables can further improve the performance of mathematical models. Fig. 2 shows predictions from January 17 to January 2023, comparing the performance of MLR, 4th degree Polynomial (P4), ANN, GRU and LSTM models. Among all the models, the best performance was observed in the GRU model with R^2 of 0.948 in testing and RMSE of 0.495 mm/min. Compared to LSTM, the GRU structure is more concise and has fewer parameters, which minimizes overfitting and improves training efficiency (Li et al., 2022). However, the performance of LSTM and ANN models did not vary much from GRU, which yielded R^2 of 0.946 and 0.944, respectively, and RMSE of 0.504 mm/min and 0.511 mm/min. In a study conducted by Chia et al. (2022), it was found that the LSTM and GRU models have tremendous potential in estimating evapotranspiration if just designed with a purpose, such as integrating an optimization algorithm. LSTM and GRU, the ANN model showed promising performance in predicting evapotranspiration in some crops like tomato (Tunali et al., 2023) and paprika (Nam et al., 2019).

The 1:1 comparison of actual transpiration with the transpiration predicted by the models is shown in Fig. 3. The improvement in model performance from MLR to GRU in predicting transpiration can be observed as the distance of the data points got closer to the 1:1 line. However, more dispersed values can be observed in the actual transpiration against transpiration calculated using the PM equation. Similar to the findings of Fernandez et al. (2010), the PM equation underestimated the actual transpiration, as shown by more predicted values that fall far below the 1:1 line. RMSE (0.598 mm/min) offers better performance in pre-
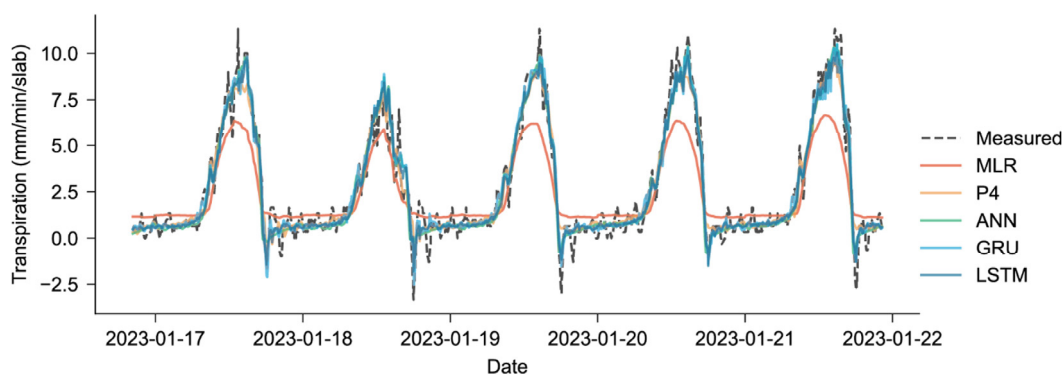

Fig. 2. Time series predictions from January 17 to January 22, 2023, comparing the performance of MLR, 4th-degree Polynomial (P4), ANN, GRU, and LSTM models.

diction compared to MLR and polynomials degrees 2 and 3. However, it has the lowest performance among all models in terms of capturing the variability in the actual transpiration at R^2 0.31. Incrocci et al. (2020) explained that PM might underestimate calculation because of the calculation of aerodynamic resistance (ra) as a function of airspeed in the original PM-Et$_o$ equation because air movement inside the greenhouses is very low in conditions of natural ventilation, much lower than in open field. To manage this condition, the use of a constant ra value of 295 s/m proposed by Fernandez et al. (2011), which is equivalent to a greenhouse air speed of 0.7 m/s, was adopted in this study. However, contrary to the result of Gallardo et al. (2016), the PM equation still underestimated the actual transpiration. Modification in the equation, such as including other significant factors such as leaf area index (LAI) in the calculation of transpiration (Jo and Shin, 2021), may be done to improve the value of aerodynamic resistance.

In terms of runtime per epoch while building deep learning models, ANN model showed the shortest average runtime per epoch which is 2.16 s, followed by LSTM and then GRU with average runtime per epoch of 47.96 s and 55.29 s, respectively. The architecture of the deep learning models is the same, but sequence lengthwas added to the LSTM and GRU model which contributed to the complexity of these models resulting to longer runtime.
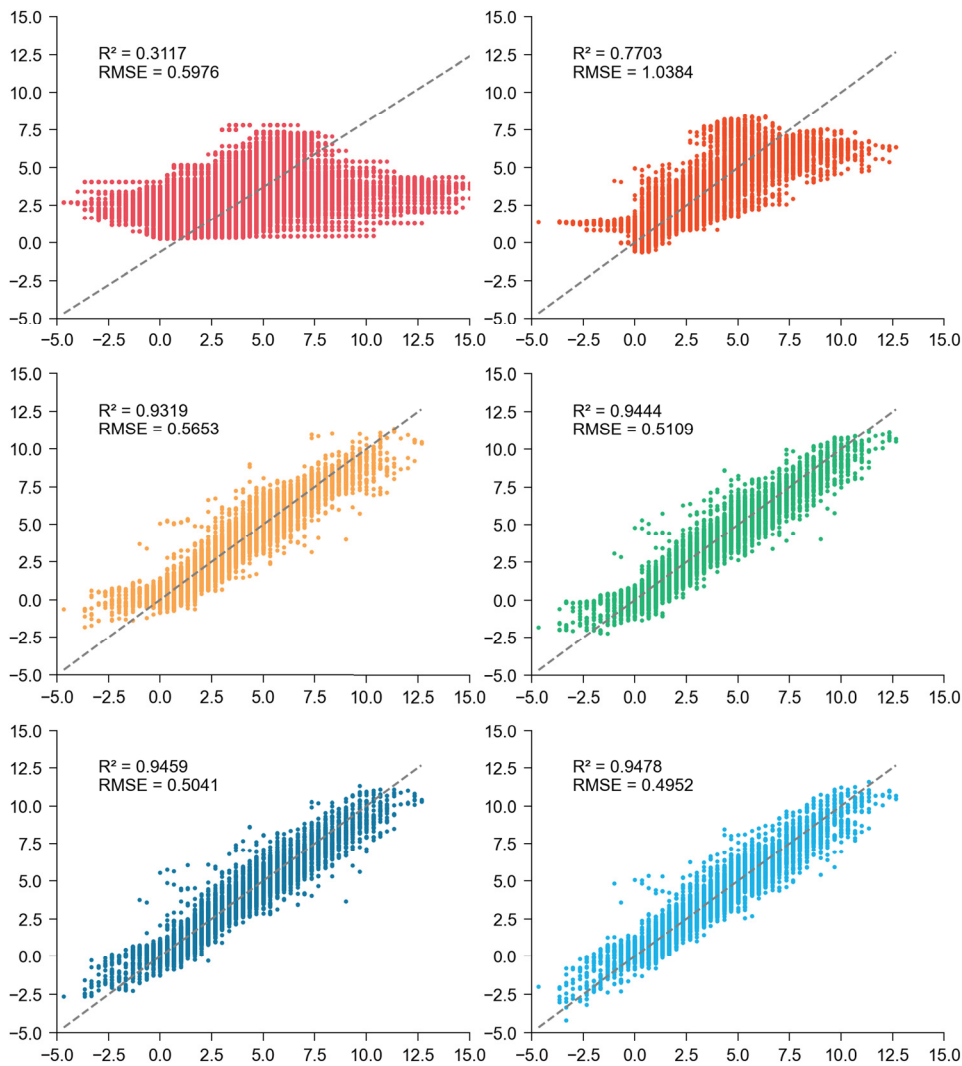
Fig. 3. Comparison between measured and estimated tomato transpiration using various models: Penman-Monteith (A), MLR (B), 4th-degree Polynomial (C), ANN (D), LSTM (E), and GRU (F).

Meanne P. Andes · Mi-young Roh · Mi Young Lim · Gyeong-Lee Choi · Jung Su Jung · Dongpil Kim

Table 3. Performance of deep learning models in prediction and in 10, 30, 60, 120, and 180-minute forecasting of transpiration.

Model	RMSE (mm/min) of Forecast				
	10-min	30-min	60-min	120-min	180-min
ANN	4.132	517376.018	-*	-*	-*
LSTM	1.036	0.919	0.930	0.755	0.634
GRU	0.577	0.402	0.489	0.697	0.953

*Value is too high in ANN forecast 60-min, 120-min, and 180-min.

The saved ANN, LSTM and GRU models were used to do forecasting for the next 10, 30, 60, 120 and 180 minutes using unseen data from May 3, 2023, to May 12, 2023. Results of forecasting are shown in Table 3.

ANN did not perform well in forecasting compared to GRU and LSTM with RMSE of 4.132 g/min and tremendously increased with increased forecasting time. GRU model showed best performance in making a 10, 30, 60, and 120-minute forecast among all deep learning models with RMSE of 0.577 mm/min. 0.402 mm/min, 0.489 g/min and 0.697 mm/min respectively. LSTM ranked second to GRU with RMSE of 1.036 mm/min, 0.919 mm/min, 0.930 mm/min and 0.755 mm/min respectively. Therefore, for this particular study, GRU is recommended for doing 10 to 120 min forecast but for longer forecasting time (180 min), LSTM is the recommended model. However, performance of deep learning models in doing forecasting should be tested using larger dataset for further verification.

Conclusion

Observed environmental variables which include inside air temperature, outside radiation and humidity were found to have highest significant correlation with transpiration among other variables hence selected as input features for model building. Inside air temperature and outside radiation are found to be positively correlated with transpiration while humidity is negatively correlated with transpiration. Mathematical models such as MLR and Polynomial regression (degrees 2, 3 and 4) modes, and deep learning models such as ANN, LSTM and GRU models were built. All models showed potential in estimating transpiration with R^2 values ranging from 0.770 to 0.948 and RMSE of 0.495 mm/min to 1.038 mm/min in testing. Deep learning models were found

to perform better in estimating transpiration compared to mathematical models. Among the deep learning models, GRU model and LSTM showed the best performance which both have 0.95 R^2 and 0.495 mm/min and 0.504 mm/min RMSE respectively. The FAO56 PM equation underestimated transpiration with RMSE of 0.598 mm/min, which is lower compared to RMSE of MLR and Polynomial degrees 2 and 3. However, it performs the least ($R^2 = 0.31$) among all models in terms of capturing the variability in actual transpiration. In terms of forecasting, GRU model performed better in doing a 10 to 120-minute forecast followed by LSTM while ANN did not perform well in doing longer time forecast. Meanwhile, LSTM performed best in forecasting longer time step. Performance of deep learning models in making forecast should still be done using larger dataset for further verification. Therefore, in this study, LSTM and GRU models are recommended for estimating transpiration of tomato in greenhouse.

Acknowledgement

Literature Cited

Allen R.G., L.S. Pereira, D. Raes, and M. Smith 1998, Crop evapotranspiration-guidelines for computing crop water requirements. FAO, Irrigation and Drainage Paper 56. Rome, Italy, p 300.

Allen R.G., W.O. Pruitt, J.L. Wright, T.A. Howell, F. Ventura, R. Snyder, D. Itenfisu, P. Steduto, J. Berengena, J.B. Yrisarry, M. Smith, L.S. Pereira, D. Raes, A. Perrier, I. Alves, I. Walter, and R. Elliot 2006, A recommendation on standardized surface resistance for hourly calculation of reference Et_o by the FAO56 Penman-Monteith method. Agric Water Manag 81:1-22. doi:10.1016/j.agwat.2005.03.007

Bejo S.K., S. Mustaffha, and W.I.W. Ismail 2014, Application

of Artificial Neural Network in predicting crop yield: A review. J Food Eng 4:1-9.

Bera D., N.D. Chatterjee, and S. Bera 2021, Comparative performance of linear regression, polynomial regression and generalized additive model for canopy cover estimation in the dry deciduous forest of West Bengal. Remote Sens Appl Soc Environ 22:100502. doi:10.1016/j.rsase.2021.100502

Chauhan N.S. 2020, Optimization algorithms in Neural Networks. KDNuggets. Available via https://www.kdnuggets.com/2020/12/optimization-algorithms-neural-networks.html Accessed August 28 2023.

Chen Z., Z. Zhu, H. Jiang, and S. Sun 2020, Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. J Hydrol 591:125286. doi:10.1016/j.jhydrol.2020.125286

Chia M.Y., Y.F. Huang, C.H. Koo, J.L. Ng, A.N. Ahmed, and A. El-Shafie 2022, Long-term forecasting of monthly mean reference evapotranspiration using Deep Neural Network: A comparison of training strategies and approaches. Appl Soft Comput 126:109221. doi:10.1016/j.asoc.2022.109221

Dahikar S., and S. Rode 2014, Agricultural crop yield prediction using artificial neural network approach. Int J Innov Res Electric Electron Instrum Control Eng 2:683-686.

De Wit C.T. 1958. Transpiration and Crop Yields. Institute of biological and chemical research of field crops and herbage. Wageningen 60:29-31. doi:10.1002/csc2.20094

Elbeltagi A., N. Kumari, J.K. Dharpure, A. Mokhtar, K. Alsafadi, M. Kumar, B. Mehdinejadiani, H.R. Etedali, Y. Brouziyne, A.R.M.T. Islam, and A. Kuriqi 2021, Prediction of combined terrestrial evapotranspiration index (CTEI) over large river basin based on machine learning approaches. Water 13:547. doi:10.3390/w13040547

Fan J., J. Zheng, L. Wu, and F. Zhang 2021, Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural network models. Agric Water Manag 245:106547. doi:10.1016/j.agwat.2020.106547

Fernandez J.E., and M.V. Cuevas 2010, Irrigation scheduling from stem diameter variations: a review. Agric For Meteorol 150:135-151. doi:10.1016/j.agrformet.2009.11.006

Fernandez M.D., S. Bonachela, F. Orgaz, R.B. Thomson, J.C. Lopez, M.R. Granados, M. Gallardo, and E. Fereres 2011, Erratum to: Measurement and estimation of plastic greenhouse reference evapotranspiration in Mediterranean climate. Irrig Sci 29:91-92. doi:10.1007/s00271-010-0210-z

Ferreira L.B., and F.F. da Cunha 2020, New Approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. Agric Water Manag 234:106113. doi:10.1016/j.agwat.2020.106113

Gallardo M., M.D. Fernandez, C. Gimenez, F.M. Padilla, and R.B. Thompson 2016, Revised VegSyst model to calculate dry matter production, critical N uptake and ETc of several vegetable species grown in Mediterranean greenhouses. Agric Syst 146:30-43. doi:10.1016/j.agsy.2016.03.014

Gallardo M., R.B. Thompson, and M.D. Fernandez 2013, Water requirements and irrigation management in Mediterranean greenhouses: the case of the southeast coast of Spain. In W Baudoin, R Nono-Womdim, N Lutaladio, A Hodder, N Castilla, C Leonardi, S de Pascale, and M. Qaryouti, eds, Good Agriculture Practices for Greenhouse Vegetable Crops: Principles of Mediterranean Climate Areas. FAO, Rome, Italy, pp 109-136.

Geelen P.A.M., J.O. Voogt, and P.A. van Weel 2020, Plant empowerment: The basic principles. Ed 2. Letsgrow.com.

Hazlett D. 2022, Importance of transpiration rates. J Glob Sci Res 10:7-8. doi:10.15651/GJARR.22.9.4

Incrocci L., R.B. Thompson, M.D.F. Fernandez, S. de Pascale, A. Pardossi, C. Stanghellini, Y. Rouphael, and M. Gallardo 2020, Irrigation management of European greenhouse vegetable crops. Agric Water Manag 242:106393. doi:10.1016/j.agwat.2020.106393

Jo W.J., and J.H. Shin 2021, Development of a transpiration model for precise tomato (*Solanum lycopersicum* L.) irrigation control under various environmental conditions. Plant Physiol Biochem 162:388-394. doi:10.1016/j.plaphy.2021.03.005

Jolliet O., and B.J. Bailey 1992, The effect if climate on tomato transpiration in greenhouses: measurements and models comparison. Agric For Meteorol 58:43-62. doi:10.1016/0168-1923(92)90110-P

Katsoulas N., and C. Stanghellini 2019, Modelling crop transpiration in greenhouses: different models for different applications. Agronomy 9:392. doi:10.3390/agronomy9070392

Li Y., J. Ye, D. Xu, G. Zhou, and H. Feng 2022, Prediction of sap flow with historical environmental factors based on deep learning technology Comput Electron Agric 202:107400. doi:10.1016/j.compag.2022.107400

Li Y., L. Gou, J. Wang, Y. Wang, D. Xu, and J. Wen 2023, An improved sap flow prediction model based on CNN-GRU-BiLSTM and factor analysis of historical environmental variables. Forests 14:1310. doi:10.3390/f14071310

Magan J.J., M. Gallardo, R.B. Thompson, and P. Lorenzo 2008, Effects of salinity on fruit yield and quality of tomato grown in soil-less culture in greenhouses in Mediterranean climatic conditions. Agric Water Manag 95:1041-1055. doi:10.1016/j.agwat.2008.03.011

Mehdizadeh S., B. Mohammadi, Q.B. Pham, and Z. Duan 2021, Development of boosted machine learning models for estimating daily reference evapotranspiration and comparison with empirical approaches. Water 13:3489. doi:10.3390/w13243489

Mohammadi B., and S. Mehdizadeh 2020, Modeling daily reference evapotranspiration via a novel approach based on

Meanne P. Andes · Mi-young Roh · Mi Young Lim · Gyeong-Lee Choi · Jung Su Jung · Dongpil Kim

support vector regression coupled with whale optimization algorithm. Agric Water Manag 237:106145. doi:10.1016/j.agwat.2020.106145

Mohammadi B., R. Moazenzadeh, K. Christian, Z., and Duan 2021, Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models. Environ Sci Pollut Res 28:65752–65768. doi:10.1007/s11356-021-15563-1

Nam D.S., T. Moon, J.W. Lee, and J.E. Son 2019, Estimating transpiration rates of hydroponically-grown paprika via an artificial neural network using aerial and root-zone environments and growth factors in greenhouses. Hortic Environ Biotechnol 60:913-923. doi:10.1007/s13580-019-00183-z

Nwogu E.C., I.S. Iwueze, and V.U. Nlebedim 2016, Some tests for seasonality in time series data. J Mod Appl Stat Methods 15:382-399. doi:10.22237/jmasm/1478002920

Ostertagová E. 2012, Modelling using polynomial regression. Procedia Eng 48:500-506. doi:10.1016/j.proeng.2012.09.545

Park Y.S., and S. Lek 2016, Chapter 7-artificial neural networks: Multilayer perceptron for ecological modeling. Dev Environ Model 28:123-140. doi:10.1016/B978-0-444-63623-2.00007-4

PASSeL (Plant and Soil Science eLibrary) 2023, Transpiration-factors affecting rates of transpiration. https://passel2.unl.edu/view/lesson/c242ac4fbaaf/6. Accessed August 28 2023

Priestly C.H.B., and R.J. Taylor 1972, On the assessment of surface heat flux and evaporation using large-scale parameters. Mon Weather Rev 100:81-92. doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Saedi R. 2022, Evaluation of multivariate regression models in estimation of evaporation and transpiration components of maize, under salinity stress conditions. Iran J Soil Water Res 53:71-84. doi:10.22059/ijswr.2022.335453.669157

Sanchez J.A., F. Rodriguez, J.L. Guzman, and M.R. Arahal 2012, Virtual sensors for designing irrigation controllers in greenhouses. Sensors 12:15244-15266. doi:10.3390/s121115244

Shao M., H.Liu, and L.Yan 2022, Estimating tomato transpiration cultivated in a sunken solar greenhouse with the Penman-Monteith, Shuttleworth-Wallace and Priestly-Taylor models in the North China plain. Agronomy 12:2382. doi:10.3390/agronomy12102382

Shin J.H., and J.E. Son 2015, Development of a real-time irrigation control system considering transpiration, substrate electrical conductivity, and drainage rate of nutrient solutions in soilless culture of paprika (*Capsicum annuum* L.). Eur J Hortic Sci 80:271-279. doi:10.17660/eJHS.2015/80.6.2

Shin J.H., J.S. Park, and J.E. Son 2014, Estimating actual transpiration rate with compensated levels of accumulated radiation for the efficient irrigation of soilless cultures of paprika plants. Agric Water Manag 135:9-18. doi:10.1016/j.agwat.2013.12.009

Shuttleworth W.J., and J.S. Wallace 1985, Evaporation from sparse crops-An energy combination theory. Q J R Meteorol 111:839-855. doi:10.1002/qj.49711146910

Tian H., P. Wang, K. Tansey, J. Zhang, S. Zhang, and H. Li 2020, An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. Agric For Meteorol 310:108629. doi:10.1016/j.agrformet.2021.108629

Tranmer M., J. Murphy, M. Elliot, and M. Pampaka 2020, Multiple linear regression (2nd Edition). Cathie March institute working paper. https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/multiple-linear-regression.pdf Accessed August 28 2023

Tu J., X. Wei, B. Huang, H. Fan, M. Jian, and W. Li 2019, Improvement of sap flow estimation by including phonological index and time-lag effect in back-propagation neural network models. Agric For Meteorol 276:107608. doi:10.1016/j.agrformet.2019.06.007

Tunali U., I.H. Tuzel, Y. Tuzel, and Y. Senol 2023, Estimation of actual crop evapotranspiration using artificial neural networks in tomato grown in closed soilless culture system. Agric Water Manag 284:108331. doi:10.1016/j.agwat.2023.108331

Vanja S., M. Eibl, and C. Hochenauer 2021, Artificial intelligence for time-efficient prediction and optimization of solid oxide fuel performances. Energy Convers Manag 230:113764. doi:10.1016/j.enconman.2020.113764.

Yildirim E., and M. Ekinci 2022, Vegetable crops: health benefits and cultivation. IntechOpen:95704. doi:10.5772/intechopen.95704

Ying X. 2019, An overview of overfitting and its solutions. J Phys Conf Ser 1168:022022. doi:10.1088/1742-6596/1168/2/022022

Yong S.L.S., J.L. Ng, Y.F. Huang, and C.K. Ang 2023, Estimation of reference crop evapotranspiration with three different machine learning models and limited meteorological variables. Agronomy 13:1048. doi:10.3390/agronomy13041048

Zarzycki K., and M. Lawrynczuk 2021, LSTM and GRU neural networks as models of dynamical processes used in predictive control: A comparison of models developed for two chemical reactors. Sensors 21:5625. doi:10.3390/s21165625

Zhang J., Y. Zhu, X. Zhang, M. Ye, and J. Yang 2018, Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J Hydrol 561:918-929. doi:10.1016/j.jhydrol.2018.04.065

Zhu Y., Z. Cheng, K. Feng, C. Zhang, C. Cao, J. Huang, H. Ye, and Y. Gao 2022. Influencing factors for transpiration rate: A numerical simulation of an individual leaf system. Therm Sci Eng Prog 27:101110. doi:10.1016/j.tsep.2021.101110

라이시미터 데이터로 학습한 수학적 및 심층 신경망 모델을 통한 온실 토마토 증산량 추정

메안 P 안데스[1] · 노미영[2] · 임미영[3] · 최경이[3] · 정정수[4] · 김동필[3]

[1]필리핀 식물산업국 국립 작물연구개발 및 생산지원센터(BPI−LBNCRDPSC) 연구원
[2]국립원예특작과학원 시설원예연구소 농업연구관
[3]국립원예특작과학원 시설원예연구소 농업연구사
[4]경상북도 농업기술원 농촌지도사

적 요. 증산은 적정 관수 관리에 중요한 역할을 하므로 수분 스트레스에 취약한 토마토와 같은 작물의 관개 수요에 대한 지식이 필요하다. 관수량을 결정하는 한 가지 방법은 증산량을 측정하는 것인데, 이는 환경이나 생육 수준의 영향을 받는다. 본 연구는 분단위 데이터를 통해 수학적 모델과 딥러닝 모델을 활용하여 토마토의 증발량을 추정하고 적합한 모델을 찾는 것을 목표로 한다. 라이시미터 데이터는 1분 간격으로 배지무게 변화를 측정함으로써 증산량을 직접 측정했다. 피어슨 상관관계는 관찰된 환경 변수가 작물 증산과 유의미한 상관관계가 있음을 보여주었다. 온실온도와 태양복사는 증산량과 양의 상관관계를 보인 반면, 상대습도는 음의 상관관계를 보였다. 다중 선형 회귀(MLR), 다항 회귀 모델, 인공 신경망(ANN), Long short-term memory(LSTM), Gated Recurrent Unit(GRU) 모델을 구축하고 정확도를 비교했다. 모든 모델은 테스트 데이터 세트에서 0.770−0.948 범위의 R^2 값과 0.495mm/min−1.038mm/min의 RMSE로 증산을 잠재적으로 추정하였다. 딥러닝 모델은 수학적 모델보다 성능이 뛰어났다. GRU는 0.948의 R^2 및 0.495mm/min의 RMSE로 테스트 데이터에서 최고의 성능을 보여주었다. LSTM과 ANN은 R^2 값이 각각 0.946과 0.944, RMSE가 각각 0.504m/min과 0.511로 그 뒤를 이었다. GRU 모델은 단기 예측에서 우수한 성능을 보였고 LSTM은 장기 예측에서 우수한 성능을 보였지만 대규모 데이터 셋을 사용한 추가 검증이 필요하다. FAO56 Penman-Monteith(PM) 방정식과 비교하여 PM은 MLR 및 다항식 모델 2차 및 3차보다 RMSE가 0.598mm/min으로 낮지만 분단위 증산의 변동성을 포착하는 데 있어 모든 모델 중에서 가장 성능이 낮다. 따라서 본 연구 결과는 온실 내 토마토 증산을 단기적으로 추정하기 위해 GRU 및 LSTM 모델을 권장한다.

추가 주제어 : 토마토 증산 모델, 회귀분석 모델, 인공신경망, LSTM, GRU