University of Birjand

Water Harvesting Research Vol. 7, No. 2, Autumn & Winter 2024, Original paper p. 234-243

10.22077/JWHR.2024.8265.1155

Canopy Temperature Estimation Using Gene Expression Programming Models and Artificial Neural Networks

Mehri Saeidinia[a*], AmirHamzeh Haghiabi[a]

[a]Department of Water Engineering, Faculty of Agriculture, Lorestan University, Khorramabad, Iran.

[*]Corresponding Author, E-mail address: saeedinia.m@lu.ac.ir

Received: 11 October 2024/ **Revised**: 29 October 2024/ **Accepted**: 13 November 2024

Abstract

Canopy temperature (T_c) is one of the essential for irrigation scheduling. Measuring canopy temperature is expensive and time-consuming. Simple approaches such as soft computing can be a good tool for this purpose because there has been no documented research in this field. In this study, the ANN (MLP with two hidden layers) and GEP models were used to estimate Tc using limited data such as the dry (T_a) and wet bulb (T_W) temperatures, saturation vapor pressure (e_s), actual vapor pressure (e_a), and the vapor-pressure deficit (VPD). Six combinations of input variables were investigated. The perfect model was selected based on statistical indices during the training and testing. Results showed that the performance of the models were influenced by the number of the input variables. The MLP models outperformed GEP models during the training and testing processes. The MLP7 (input variables: e_s and e_a) with MSE of 1.08 $^{\circ}$C, RMSE of 1.04 $^{\circ}$C, and R^2 of 0.92 in the training phase and MSE of 1.02, RMSE of 1.00, and R^2 of 0.95 in the validation phase was selected as the perfect model among MLP models. The GEP11(input variables: T_a, T_W, e_s, e_a, and VPD) with MSE of 1.32, RMSE of 1.15, and R^2 of 0.89 in the training phase and MSE of 0.91, RMSE of 0.95, and R^2 of 0.95 in the validation phase was also the perfect model among GEP models. Accordingly, the proposed GEP and MLP models can be drawn on as a perfect model for estimating T_C.

Keywords: ANN, Climate data, GEP, Irrigation scheduling, MLP.

1. Introduction

Canopy temperature is one of the main parts of the soil-plant-atmosphere energy that can be used as a valuable tool for showing the water status of plants and irrigation scheduling (Sánchez-Piñero et al., 2022). In addition, leaf temperature influences photosynthesis, respiration, and transpiration processes (Blonder and Michaletz, 2018). Canopy temperature can be affected by climatical factors such as air temperature, wet-bulb temperature, actual vapor pressure, saturation vapor pressure, and vapor-pressure deficit (O'shaughnessy et al., 2011, Blonder and Michaletz, 2018). Measuring canopy temperature with an infrared thermometer in an open field is time-consuming, high-priced, and unaffordable. Therefore, it is essential to estimate canopy temperature without cost using straightforward approaches such as data mining techniques (Van Klompenburg et al., 2020). Various data mining techniques are being used extensively for many purposes in the agricultural sector, such as evapotranspiration (Antonopoulos and Antonopoulos, 2017; Valipour et al., 2019), crop yields (Taherei Ghazvinei et al., 2018), soil temperature (Seifi et al., 2021), and leaf area (Küçükönder et al., 2016).

Artificial neural networks (often referred to simply as neural networks or connectionist models) offer a solution for tackling complex pattern-oriented challenges related to categorization and time-series analysis. Their nonparametric characteristic allows these models to be created without needing prior knowledge of the data population's distribution or potential interactions among variables,

which is a requirement for traditional parametric statistical methods (Walczak, 2019).

Artificial neural network (ANN), based on a deep learning technique, is increasingly applied in nonlinear data modeling (Adisa et al., 2019, Mahanti et al., 2022). Many researchers have used ANNs for simulating different parameters in agriculture such as Heramb et al. (2022), Monteiro et al. (2021), Han et al. (2021), Gavahi et al. (2021) and Elbeltagi et al. (2020).

Gene expression programming (GEP) programs consist of intricate tree structures that evolve and adjust by altering their sizes, shapes, and components, similar to a living organism. Additionally, akin to living beings, GEP's computer programs are encoded in straightforward linear chromosomes of a predetermined length (Ferreira, 2001). These software, a genotype/phenotype genetic algorithm, is an artificial intelligence model widely used to model nonlinear operations. According to the literature, these models are accurate and can be developed and evaluated for different purposes.

Maize is one of the main crops in Iran. Hence, the canopy temperature estimation for this plant can be a cost-effective tool for water stress management and irrigation schedule. Best of our knowledge, only a little research has been conducted on the canopy temperature estimation. Thus, this study aimed to develop a new approach for estimating the canopy temperature of Maize using GEP and ANN models (MLP) and evaluating the performance of these models using statistical indices.

2. Materials and Methods

2.1. Experiment area

In this study, two experiments were conducted using summer and winter planting of Maize during 2013-2014 at the research field of Shahid Chamran University of Ahvaz, Ahvaz, Khuzestan province, Iran (31˚18'10"N, 48˚39'41"E, 20m above sea level (Figure1)). The soil of the area was silty loam. In the first and second experiments, grain Maize (variety SC704) was planted on July 23, 2013, and February 23, 2014. Surface irrigation with seven days interval was applied during two growing seasons. The soil-water status was monitored and measured before irrigation using the gravimetric method to determine the water required. The average of maximum temperature, minimum temperature, relative humidity, wind speed, and precipitation during the first growing season was 34.94˚C, 22.25˚C, 41%, 5.42 $m.s^{-1}$, and 1.9 $mm.d^{-1}$. These values were 38.2˚C, 19.35˚C, 39.37%, 5.55 $m.s^{-1}$, and 2.5 $mm.d^{-1}$ for the second season. The field capacity and wilting point in the upper 90 cm soil profile were obtained at 0.32 and 0.15 m^3m^{-3}.

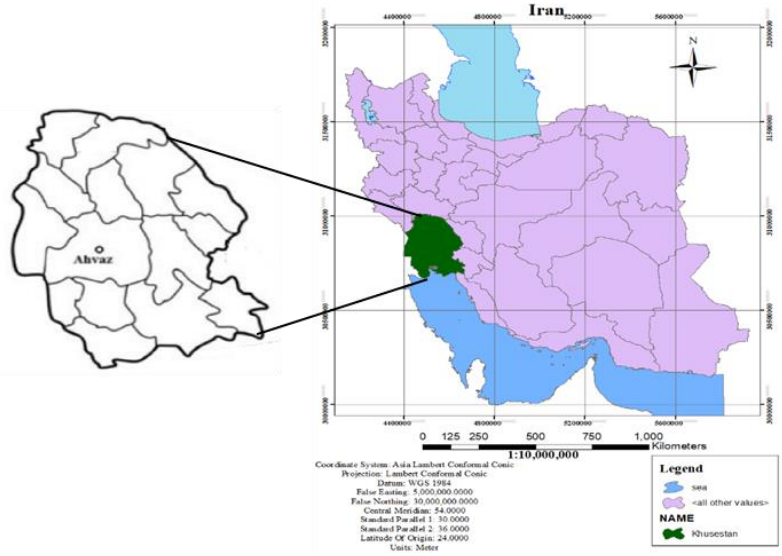

Fig. 1. Location map of the study area

2.2. Field measurements

Canopy temperature measurements were taken from 27 August 2013 (first growing season) and 10 April 2014 (second growing season), respectively when canopy cover was approximately 80-85%. The measurement was

taken during 8 days in the first season and 7 days in the second season (Table 1) using a handheld infrared thermometer (IRT) equipped with an 8-14µm band-pass spectral filter with a minimum measurement diameter of 6 mm and a distance-to-spot size ratio (D:S) of 8:1 (Reyteck model, Fluke Corporation, USA). Measuring through the days following irrigation, started from 08.00 am to 02.00 pm on an hourly interval under clear sky from the north, south, east, and west directions, and then averaged. The dry and wet bulb temperatures were also calculated at each time of measurement using an adjusted psychrometer at the height of 2 m near the research field.

The saturation vapor pressure (e_s), the actual vapor pressure (e_a), and the vapor pressure deficit (VPD) of the air were measured as (Allen et al., 1998):

$$e_s = 6.047 \times exp[17.27\frac{T_{dry}}{T_{dry}+237.3}] \quad (1)$$

$$e_a = \left[\begin{matrix}\left(0.6108 \times \exp\left(17.27\frac{T_{dry}}{T_{dry}+237.3}\right)\right) \\ -\left(0.0012 \times 101.03 \times T_{dry} - T_{wet}\right)\end{matrix}\right] \times 9.9\ 01 \quad (2)$$

$$VPD = e_s - e_a \quad (3)$$

where e_s is the saturation vapor pressure (mbar), T_{dry} is the air temperature (˚C), T_{wet} is the wet bulb temperature, e_a is the actual vapor pressure (mbar), and VPD is the vapor-pressure deficit (mbar). Statistical properties of the data used in this research are provided in Table 2.

The correlation of the measured meteorological variables with T_c is depicted in Figure 2. It can be seen that the linear correlation between the canopy temperature and T_a (0.835), T_w (0.850), and e_s (0.773) is stronger than that of e_a (0.550) and VPD (0.363). It is essential to evaluate the combination of these variables to estimate Tc.

2.3. Estimation of T_c using ANN and GEP

2.3.1. ANN models

In this study, MLP was applied to predict T_c using a different combination of variables (Table 4), such as T_a (˚C), T_w (˚C), e_s (mbar), e_a (mbar), and VPD (mbar). After normalizing, all data (100 data points) were randomly divided into three parts, including training (70%), testing (15%), and validation (15%). In this study, MATLAB 16[b] was used to analyze various trials to find the best MLP structure. To analyze various architectures of MLP models, feed-forward networks with four layers, including the input layer, two hidden layers, and the output layer were used. Levenberg-Marquart backpropagation was utilized for the training algorithm. To find the best number of neurons in hidden layers, a trial and error was conducted. In this process, the number of neurons in the hidden layer varied from 2 to 10.

Table 1. The dates of measurements in two growing seasons

First season (Summer Maize)	Second season (Winter Maize)
2013-08-27	2014-04-10
2013-09-03	2014-04-17
2013-09-10	2014-04-24
2013-09-17	2014-05-01
2013-09-24	2014-05-08
2013-10-03	2014-05-21
2013-10-10	2014-05-29
2013-10-17	

Table 2. Statistical properties of the study data

Variables	Minimum	Maximum	Average	Median	Standard deviation
T_a (˚C)	17.5	46	33.745	34	6.048
T_w (˚C)	14	37	26.14	26	4.757
e_s (mbar)	19.802	99.849	54.449	52.67	17.715
e_a (mbar)	11.026	51.62	25.588	22.096	9.862
VPD (mbar)	5.196	61.227	28.875	27.136	13.755
T_c (˚C)	18.703	37.921	29.969	30.108	3.748

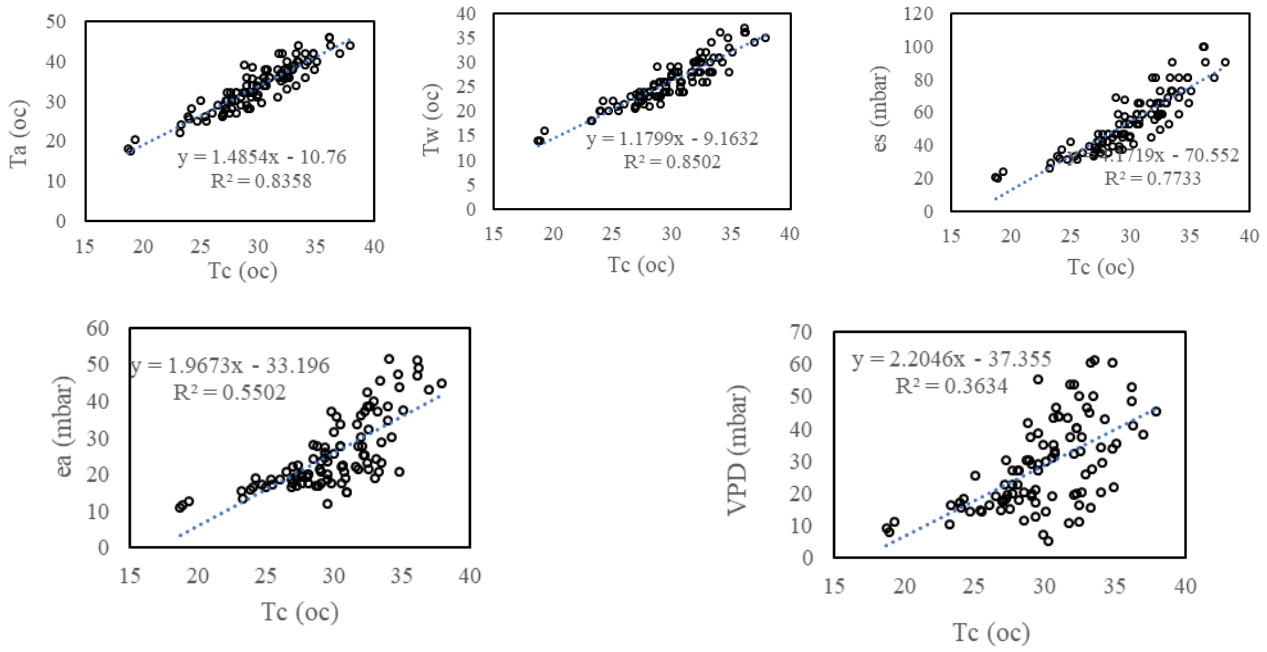

Fig. 2. Correlation of the measured meteorological variables with T_c

2.3.2. GEP models

GEP is a genotype/phenotype genetic algorithm that uses a population of individuals (chromosomes). In the GEP, chromosomes that consist of one or more genes are encoded as linear entities of fixed length. Then, these chromosomes are converted into nonlinear entities of various sizes and shapes (expression trees). The individuals are evaluated using fitness functions. The solving process starts with creating a random individual generation as the initial population.

Then, these individuals are assessed by fitness functions. Afterward, the selected generation is reproduced by modification processes such as mutation, inversion, transposition, and recombination. These processes are repeated to find the best solution. GeneXproTools 5.0 was used to conduct GEP modeling. The result of GEP modeling was a clear mathematical expression that indicates the relationship between independent variables (inputs) and dependent outputs. In this study, the input variables for GEP models were different combinations of T_a, T_w, es, ea, and VPD.

Seventy percent of all data (70 data points) were considered for the training, 15% (15 data points) for testing phase, and the rest of the data were used for validation (15 data points). T_c (˚C) was used as the output variable.

2.4. Statistical analysis

In this study, the performance of developed models was evaluated using statistical indices such as the coefficient of determination (R^2), the mean some of the square (MSE), and the root means square error (RMSE).

R^2 indicates the correlation between the estimated values of the models and the measured values. It ranges from 0 to 1, where 1 shows the perfect correlation. RMSE and MSE range from 0 to +∞, and the perfect value is 0.

$$R^2 = \frac{(\sum_{i=1}^{n}(T_{c_{o,i}} - \overline{T_{c_o}})(T_{c_{e,i}} - \overline{T_{c_e}}))^2}{\sum_{i=1}^{n}(K_{c_{o,i}} - \overline{K_{c_o}})^2 . \sum_{i=1}^{n}(K_{c_{e,i}} - \overline{K_{c_e}})^2} \quad (4)$$

$$MSE = \frac{\sum_{i=1}^{n}(T_{c_{o,i}} - T_{c_{e,i}})^2}{n} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(T_{c_{o,i}} - T_{c_{e,i}})^2}{n}} \quad (6)$$

where $T_{c_{o,i}}$ and $T_{c_{e,i}}$ are the observed (measured) and the estimated values of T_c. $\overline{T_{c_o}}$ and $\overline{T_{c_e}}$ indicate the mean values of the observed and estimated values. Note that in this study, RMSE and MAE are in ˚C.

3. Results and Discussion

3.1. ANN models with two hidden layers and different input variables

11 top MLP models were obtained for 11 combinations of input data variables based on the MSE and RMSE indices. To find the best number of neurons for two hidden layers, trial and error was conducted.

The values of MSE, RMSE, and R^2 for each model are indicated in Table 3. MLP1 structure with the corresponding weights and bias values are illustrated in Figure 3.

As shown in Table 3, 11 developed models were divided into five main groups to examine the effect of different input values in MLP models to estimate TC values using limited data. It can be found from the Table 3 that the performance of each model during training and testing was approximately the same. A comparison between the result of different groups in Table 3 showed that, in group one, MLP2, MLP3, and MLP1 with RMSE of 1.32, 1.41 and, 1.41°C in the training phase and 1.33, 1.38 and, 1.4°C in the testing phase respectively, have better performance than MLP5 and MLP4 with the RMSE of 2.58 and 2.24 °C in the training phase and 2.55 and 2.28 °C in the testing phase. These results are also confirmed by the MSE and R^2 values. It can be found that T_W, e_s, and e_a can be proposed as better input variables of MLP models if one input variable is used for estimating TC.

In group 2, MLP6 which applied Ta and T_w as input variables, the MSE, and RMSE decreased by 44% and 24.8 % during the training and 45.6% and 26.4% during the testing compared to MLP1 which used only Ta.

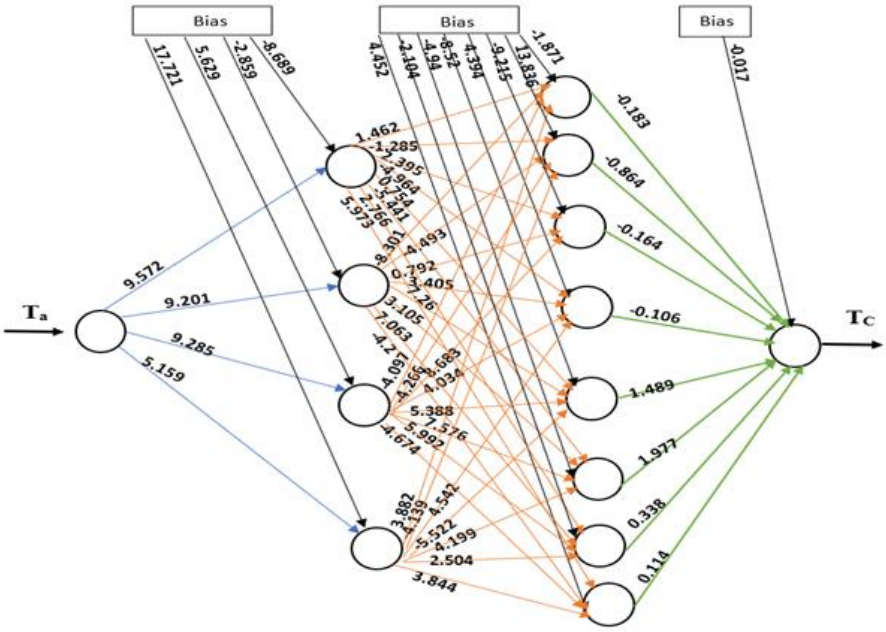

Fig. 3. The best MLP6 structure

Table 3. The best number of hidden neurons, activation functions and structures of MLP models

Model	Input Variable combinations	Group	structure	Training`			Testing		
				MSE (°C)	RMSE (°C)	R^2	MSE (°C)	RMSE (°C)	R^2
MLP1	T_a	1	1-4-8-1	2	1.41	0.87	1.95	1.4	0.85
MLP2	T_w	1	1-2-6-1	1.74	1.32	0.88	1.76	1.33	0.83
MLP3	e_s	1	1-2-3-1	1.99	1.41	0.87	1.92	1.38	0.81
MLP4	e_a	1	1-6-5-1	5.02	2.24	0.65	5.2	2.28	0.78
MLP5	VPD	1	1-6-7-1	6.67	2.58	0.50	6.52	2.55	0.65
MLP6	T_a+ T_w	2	2-5-4-1	1.12	1.06	0.91	1.06	1.03	0.91
MLP7	e_s+e_a	2	2-3-6-1	1.08	1.04	0.92	1.02	1.00	0.95
MLP8	T_a+ T_w+VPD	3	3-2-3-1	1.31	1.14	0.91	1.27	1.13	0.79
MLP9	e_s+e_a +VPD	3	3-3-3-1	1.22	1.10	0.91	1.20	1.09	0.93
MLP10	T_a+ T_w+ e_s+e_a	4	4-4-3-1	1.12	1.06	0.92	1.14	1.07	0.88
MLP11	T_a+ T_w+ e_s+e_a+VPD	5	5-3-2-1	1.34	1.16	0.89	1.30	1.14	0.92

Levenberg-Marquardt was used as train function. Levenberg-Marquardt is a popular trust region algorithm used in computer science to minimize a function by internally modeling a trusted region with a quadratic function. It is sensitive to initial starting parameters and is commonly implemented using finite differences to approximate the Jacobian matrix (Chakrabarti et al., 2012; Zeynali and Hashemi, 2016).

In another case of input variable combinations of group2, the RMSE value of

MLP7, which used ea and es as the input variable, decreased by 54% in the training phase and 56% in the testing phase compared to MLP4 which used only ea as an input variable. A comparison from the statistical indices of models in group 1 and group 2 shows that MLP models with more than one input variable show better performance. This reflects that of Heramb et al. (2022), who also found that the least RMSE is obtained when one variable is used to estimate evapotranspiration using GEP and ANNs models.

Leaf temperature could not be isolated as a function of one parameter such as air temperature. It is a term for nonlinear functions containing vapor pressure, convective resistance, air temperature, mas density of air, wind speed, and stomatal ratio, etc. (Blonder and Michaletz, 2018). Mostafa et al. (2012) and SAMMEN (2013) also showed that combining all parameters performs better than using individual parameters.

In group 3, MLP8 (with a combination of Ta, Tw and VPD), and MLP9 (with a combination of es, ea and VPD) were examined. The results showed that adding this parameter had no significant influence on estimating Tc because the values of MSE in MLP8 and MLP 9 roughly increased (7.4% and 13%) compared to MLP6 and MLP7.

The integration of input variables of MLP6 (Ta, Tw), and MLP7 (es, ea) were used in MLP10 to examine the influence of using four input variables to estimate TC (group 4). The result showed that this integration did not improve the estimation process significantly because the MSE value in MLP10 during testing increased by 7.5% and 11.8 % compared to MLP6 and MLP7.

MLP11 in group 5, which used all the input variables, achieved MSE of 1.34 and 1.30 ^{0}C and the RMSE of 1.16 and 1.14 ^{0}C during the training and testing. A comparison of the MSE and RMSE of MLP11 with those of MLP6 and MLP7 shows that, this combination also couldn't improve the estimation of TC. Thus, the models that showed better results are briefly sorted according to the MSE, RMSE and, R^2 values as follows: MLP7> MLP6 >MLP10> MLP9> MLP8> MLP11> MLP2> MLP3> MLP1> MLP4> MLP5.

3.2. Developed GEP models

The effective parameters utilized in GEP models are illustrated in Table 4. Mathematical operators used to extract the best result include basic arithmetic operators (+, -, *, /) and mathematical functions (√, ln(x), X^2, X^3) (Shiri, 2017). The linking function "addiction" was chosen to link the mathematical terms (Mattar, 2018). In this study, three genes were applied in all models. MSE and RMSE were used for the fitness function. Other parameters in Table 4, such as the mutation rate, the transposition rate, and the recombination rate were considered constant in all models. The best-extracted equations (an arithmetic form of expression trees) of all evaluated models are given in Table 5.

The performance of the models during training and testing was evaluated using statistical indices (Table 6). The Taylor Diagram which shows the R^2 values of GEP models during testing was indicated in Figure 3.

As shown in Table 6 and Figure 4, during training, the MSE, RMSE, and R^2 values of the developed models ranged from 1.29 to 7.3 °C, 1.13 to 2.7°C, and 0.45 to 0.91, respectively. These values during testing ranged from 0.91 to 7.33 °C, 0.95 to 2.7°C and 0.65 to 0.95, respectively.

Results showed that in the group one, GEP2, GEP3, and GEP1 with RMSE of 1.27, 1.46 and, 1.5°C during training and 1.35, 1.37 and, 1.38°C during testing, have better performance than GEP5 and GEP4 with the RMSE of 2.7 and 2.27 °C during training and 2.7 and 2.17 °C during testing. Looking at Fig.5, it is apparent that GEP4 and GEP5 showed a weak correlation with measured T_C.

Models of Group 2 that used two variables as input, consist of GEP6 and GEP7. In GEP6, which used Ta and T_w, the MSE decreased by 40% and 17.2 % during training and 42.1% and 40% during testing compared to GEP1 and GEP2, which used only T_a and T_W, respectively. The MSE value of GEP7 which used es and ea as the input variable, decreased by 30% and 71% during training and by 45.5% and 78.3% during testing compared to GEP3 and GEP4. A comparison of the statistical indices of models in group 1 and group 2 revealed that using two input variables

also increased the ability of GEP models to estimate T_C.

In group 3, the MSE of GEP8, which used Ta, Tw, and VPD, decreased by 3.7% during training compared to GEP6, and the MSE of GEP9 increased by 9.3% compared to GEP7. Comparing the R^2, RMSE and MSE values, it can be illustrated that no significant improvement in estimating T_C was detected by using VPD. These results are consistent with our earlier finding, which showed that in MLP models, VPD was not an influential variable for estimating T_C.

GEP10, which applied T_a, T_w, e_s, and e_a as input variables, achieved the MSE of 1.33 °C during training, which was not significantly different with those of GEP6 and GEP7. Although the results showed that GEP10 trained roughly better than GEP7, the derivatives of the MSE during testing indicated that the performance of GEP10 (1.09) was not improved than GEP7 (1.03). These results are in line with those of earlier finding about comparing MLP10 with MLP6 and MLP7.

Table 4. Effective parameters during optimal evolution in GEP for T_c modeling

Parameters	Values
Number of chromosomes	30
Head size	8
Number of genes	3
Independent variables	T_a, T_w, e_s, e_a, VPD
Dependent output	T_c
Mathematical operators	+, -, *, /, √, ln(x), X^2, X^3
Fitness function	MSE
Linking functions	Addition
Mutation rate	0.00138
Inversion rate	0.00546
Gene transposition rate	0.00277
IS transposition rate	0.00546
RIS transposition rate	0.00546
Gene recombination rate	0.00277
One-point recombination rate	0.00277
Two-point recombination rate	0.00277

Table 5. The best equations extracted after training and testing

Model	Extracted equation
GEP1	$\mathrm{TC} = \ln\left(\left(-289.29 + {T_a}^2\right) \times T_a\right) + (0.382T_a + 0.715) + \sqrt{6.5 \times \sqrt{T_a}}$
GEP2	$TC = \left(1 + \sqrt{35.36T_W}\right) + \left(\frac{\frac{2.64}{T_W} - 13.49}{\frac{T_W}{2.64} + 2.64}\right)^3 - \left(\frac{1}{2T_W - 57.4}\right)$
GEP3	$TC = \left(\sqrt{{e_s}^2 + 3.78} + \frac{e_s - 0.68}{0.68e_s}\right) + \left((-5.87 - e_s) + (10.16 \ln e_s)\right) + \left(\left(\frac{e_s + \frac{e_s}{4.65}}{14.89 - e_s}\right) - (\ln e_s - 1.72)\right)$
GEP4	$TC = \left(\left(\sqrt{1.9e_a}\right) + (8.71 - e_a)\right) + (\ln(\ln(e_a - 10))) + \left(\sqrt{4e_a} \ln 2e_a\right)$
GEP5	$TC = \left(\ln\left(\left(\left(\frac{123.9}{VPD - 5.54}\right) - (VPD^2)\right)^2\right)\right) + \left(\left(\frac{-2.31}{VPD - 7.58}\right) - VPD\right) + (VPD + 1725)$
GEP6	$TC = \left(T_W\left(\ln\left(\sqrt{\ln\left(42.45 - \frac{T_a}{7.28}\right)}\right)\right)\right) + \left(\ln\left(-7.06\left((10.94 - T_W) + (4.21/T_W)\right)\right)^2\right) + \left(\left(\frac{{T_a}^2}{T_W}\right) \times \left(\frac{T_a + T_W}{{T_W}^2}\right)\right)$
GEP7	$TC = \left(e_a + \left(\frac{15.03}{7.96 + (7.93 - e_s)}\right)\right) + \left(\left(\sqrt{e_s} + 6.1\right) - (e_a - 9.31)\right) + \left(\sqrt{(e_s - 29.44) + (2e_a - 10.7)}\right)$
GEP8	$TC = \left(\frac{{T_W}^3}{56.92T_a\sqrt{T_a}}\right) + \left(\sqrt{\left(2T_a\sqrt{T_a}\right) - (-9.91T_a - T_W)}\right) + \left(\frac{(T_W - 23.33) + (T_W - 4.6)}{VPD}\right)$
GEP9	$TC = \left((\ln(51.67 - e_a)) + \left(\left(\frac{e_a}{2.49}\right) + (VPD + 6.82)\right)\right) + \left(\frac{(-3.69e_s) + (VPD + 7.83)}{0.59(-6.34 - e_a)}\right) + \left(-1.05\left(\left((VPD - 0.68) + \left(\frac{20.57}{VPD}\right)\right) - \sqrt{5.6}\right)\right)$
GEP10	$TC = T_W + \left(\sqrt{\left(\frac{2T_W}{e_s + 5.31}\right)^4} + 4.11\right) + \left(\frac{(\ln e_s) - \left((T_W - e_a) - \left(\frac{T_W}{T_a}\right)\right)}{-3.61}\right)$
GEP11	$TC = \left(\ln\left(52.77 + (VPD - e_s)\right) + 16\right) + \left(\left(T_W + \left(\frac{e_s}{2e_a}\right)\right) - \left(\sqrt{7.77 + e_a}\right)\right)\left((e_a + 7.11) + \left(-18 - (e_s - VPD)\right)\right)$

Table 6. Statistical indices of GEP models during the training and testing phase

Model	Input Variable combinations	Group	Training`			Testing		
			MSE (°C)	RMSE (°C)	R^2	MSE (°C)	RMSE (°C)	R^2
GEP1	T_a	1	2.24	1.5	0.85	1.9	1.38	0.87
GEP 2	T_w	1	1.62	1.27	0.89	1.84	1.35	0.81
GEP 3	e_s	1	2.14	1.46	0.86	189	1.37	0.82
GEP 4	e_a	1	5.17	2.27	0.63	4.74	2.17	0.78
GEP 5	VPD	1	7.30	2.7	0.45	7.33	2.7	0.65
GEP 6	T_a+ T_w	2	1.34	1.16	0.91	1.1	1.2	0.89
GEP 7	e_s+e_a	2	1.5	1.23	0.88	1.03	1.0.2	0.95
GEP 8	T_a+ T_w+VPD	3	1.29	1.13	0.91	2.03	1.42	0.71
GEP 9	e_s+e_a +VPD	3	1.64	1.28	0.87	1.3	1.14	0.91
GEP 10	T_a+ T_w+ e_s+e_a	4	1.33	1.15	0.91	1.09	1.04	0.87
GEP 11	T_a+ T_w+ e_s+e_a+VPD	5	1.32	1.15	0.89	0.91	0.95	0.95

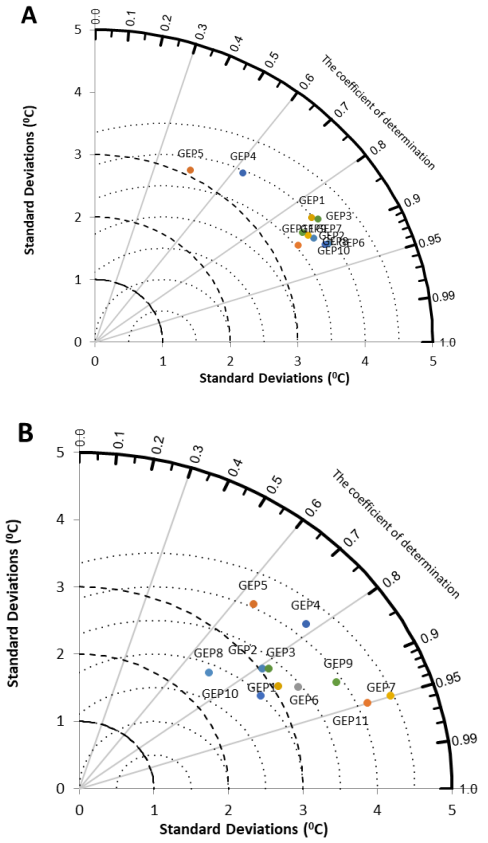

Fig. 4. The determination coefficient of GEP models during training (A) and testing (B)

All of the meteorological variables were implemented in GEP11. This model achieved the MSE of 1.33 ^{0}C and 0.91 ^{0}C during training and testing. Comparison of these results with those of GEP10 confirmed that the MSE and RMSE values in testing decreased by 16.5%, and 8.6% and the R^2 value increased by 8.7%, and ranked as the first GEP model. The models that showed better results are briefly sorted according to the MSE, RMSE and R^2 values in testing as follows: GEP11>GEP7> GEP6≥ GEP10> GEP9> GEP8> GEP 2> GEP 3> GEP 1> GEP4> GEP5. A comparison between statistical indices of MLP and GEP models in showed that MLP models trained and performed better than GEP models.

3.3. Validation of developed GEP and ANN models

In this study, 15% of the data was used for validation tests. The statistical indices for evaluated models are shown in Table 7. Mahanti et al. (2022) reported that the best model could be chosen based on a validation tests. Results of the validation test showed that MLP models performed better than GEP models to estimate T_C. MLP and GEP models with multiple input variables achieved better statistical indices. The data in Table 7 suggest that the perfect MSE and RMSE were observed in MLP7 and GEP7, and the worst values were observed in MLP5 and GEP5. Although GEP11 performed better than GEP7 in training and testing, the result was versus in validation. On the other hand, GEP7 performed better than GEP11. It seems it is because of the different statistical properties in data sets used in various processes. Maier et al. (2010) reported that the best possible model would obtain if the statistical properties of the training, testing, and validation data sets were the same.

4. Conclusion

This study aimed to evaluate the ability of ANN (MLP) and GEP models to estimate the canopy temperature (T_c) using limited climate data. To this purpose, various combinations of climatic data such as T_a, T_w, e_s, e_a, and VPD were used as input variables for the models. After training and testing, they were evaluated using statistical indices. Results suggest that the performance of the models is influenced by the number of input variables. In most cases, the MLP models outperformed the GEP ones

for estimating T_c, and the best models introduced in this study were MLP7, which used es and ea as input variables and GEP11 which used all the input variables.

The results confirmed that GEP and ANN models could be used for irrigation scheduling. It is notable that, so far, there has been no comprehensive research on predicting leaf temperature using meteorological parameters. Choosing between GEP and ANN (MLP) depends on the problem context, the nature of the data, and the desired outcome. Neural networks possess strong abilities to manage intricate, high-dimensional data, adjust to new information, and reveal underlying patterns. However, these strengths are accompanied by challenges, such as the requirement for extensive datasets, difficulties in interpretability, and substantial computational demands.

Table 7. Statistical indices of various ANN models in the validation phase

Model	MSE (°C)	RMSE (°C)	R^2	Model	MSE (°C)	RMSE (°C)	R^2
MLP1	1.95	1.40	0.78	GEP1	1.45	1.2	0.84
MLP2	1.79	1.34	0.88	GEP2	1.29	1.13	0.93
MLP3	1.92	1.4	0.81	GEP3	1.78	1.33	0.82
MLP4	5.8	2.4	0.34	GEP4	6.69	2.58	0.22
MLP5	6.44	2.53	0.36	GEP5	7.8	2.79	0.23
MLP6	1.14	1.07	0.94	GEP6	1.56	1.25	0.85
MLP7	1.02	1.01	0.95	GEP7	1.3	1.15	0.91
MLP8	1.25	1.11	0.94	GEP8	1.8	1.34	0.91
MLP9	1.25	1.11	0.92	GEP9	1.58	1.26	0.91
MLP10	1.1	1.05	0.93	GEP10	1.57	1.25	0.91
MLP11	1.3	1.14	0.93	GEP11	1.54	1.24	0.92

A thorough understanding and management of these pros and cons are essential for effectively utilizing neural networks across different applications (Khalilov et al., 2021). In contrast to formulations based on artificial neural networks (ANN), which can be overly complex for practical use, models derived from Gene Expression Programming (GEP) offer estimation equations that are relatively simple and suitable for practical design and even manual calculations (Kontoni et al., 2022).

5. Disclosure statement

No potential conflict of interest was reported by the authors.

6. References

Adisa, O. M., Botai, J. O., Adeola, A. M., Hassen, A., Botai, C. M., Darkey, D. & Tesfamariam, E. (2019). Application of artificial neural network for predicting maize production in south africa. *Sustainability, 11*, 1145.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. Fao, rome, 300, d05109.

Antonopoulos, V. Z. & Antonopoulos, A. V. (2017). Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Computers and electronics in agriculture, 132*, 86-96.

Blonder, B. & Michaletz, S. T. (2018). A model for leaf temperature decoupling from air temperature. *Agricultural and forest meteorology, 262*, 354-360.

Chakrabarti, G., Grover, V., Aarts, B., Kong, X., Kudlur, M., Lin, Y., ... & Wang, J. Z. (2012). CUDA: Compiling and optimizing for a GPU platform. *Procedia Computer Science, 9*, 1910-1919.

Elbeltagi, A., Zhang, l., Deng, J., Juma, A. & Wang, K. (2020). Modeling monthly crop coefficients of maize based on limited meteorological data: a case study in nile delta, egypt. *Computers and electronics in agriculture, 173*, 105368.

Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*.

Gavahi, K., Abbaszadeh, P. & Moradkhani, H. (2021). Deepyield: a combined convolutional neural network with long short-term memory for crop yield forecasting. *Expert systems with applications, 184*, 115511.

Han, X., Wei, Z., Zhang, B., Li, Y., Du, T. & Chen, H. (2021). Crop evapotranspiration prediction by considering dynamic change of crop coefficient and the precipitation effect in back-propagation neural network model. *Journal of hydrology, 596*, 126104.

Heramb, P., Singh, P. K., Rao, K. R., & Subeesh, A. (2023). Modelling reference

evapotranspiration using gene expression programming and artificial neural network at Pantnagar, India. I*nformation processing in agriculture, 10*(4), 547-563.

Khalilov, D. A., Jumaboyeva, N. A. K., & Kurbonova, T. M. K. (2021). Advantages and Applications of Neural Networks. *Academic research in educational sciences*, *2*(2), 1153-1159.

Kontoni, D. P. N., Onyelowe, K. C., Ebid, A. M., Jahangir, H., Rezazadeh Eidgahee, D., Soleymani, A., & Ikpa, C. (2022). Gene expression programming (GEP) modelling of sustainable building materials including mineral admixtures for novel solutions. *Mining*, *2*(4), 629-653.

Küçükönder, H., Boyaci, S. & Akyüz, A. (2016). A modeling study with an artificial neural network: developing estimationmodels for the tomato plant leaf area. T*urkish journal of agriculture and forestry, 40*, 203-212.

Mahanti, N. K., Upendar, K. & Chakraborty, S. K. (2022). Comparison of artificial neural network and linear regression model for the leaf morphology of fenugreek (trigonella foenum graecum) grown under different nitrogen fertilizer doses. *Smart agricultural technology*, *2*, 100058.

Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental modelling & software, 25*, 891-909.

Mattar, M. A. (2018). Using gene expression programming in monthly reference evapotranspiration modeling: a case study in egypt. *Agricultural water management, 198*, 28-38.

Monteiro, A. L., de Freitas Souza, M., Lins, H. A., da Silva Teófilo, T. M., Júnior, A. P. B., Silva, D. V., & Mendonça, V. (2021). A new alternative to determine weed control in agricultural systems based on artificial neural networks (ANNs). *Field Crops Research, 263*, 108075.

Mostafa, A. B., Thamer, A. M., Abdul, H. G. & Mohd, A. M. S. (2012). Prediction of evaporation in tropical climate using artificial neural network and climate based models. *Scientific research and essays, 7*, 3133-3148.

O'shaughnessy, S., Evett, S., Colaizzi, P. & Howell, T. (2011). Using radiation thermography and thermometry to evaluate crop water stress in soybean and cotton. *Agricultural water management, 98*, 1523-1535.

Sammen, S. S. (2013). Forecasting of evaporation from hemren reservoir by using artificial neural network. *Diyala journal of engineering sciences, 6*, 38-33.

Sánchez-Piñero, M., Martín-Palomo, M., Andreu, L., Moriana, A. & Corell, M. (2022). Evaluation of a simplified methodology to estimate the cwsi in olive orchards. *Agricultural water management, 269*, 107729.

Seifi, A., Ehteram, M., Nayebloei, F., Soroush, F., Gharabaghi, B. & Torabi Haghighi, A. (2021). Glue uncertainty analysis of hybrid models for predicting hourly soil temperature and application wavelet coherence analysis for correlation with meteorological variables. *Soft computing, 25*, 10723-10748.

Shiri, J. (2017). Evaluation of fao56-pm, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of iran. *Agricultural water management, 188*, 101-114.

Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K. B. W., Alizamir, M., Shamshirband, S. & Chau, K.W.(2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. E*ngineering applications of computational fluid mechanics, 12*, 738-749.

Valipour, M., Gholami Sefidkouhi, M. A., Raeini-Sarjaz, M. & Guzman, S. M. (2019). A hybrid data-driven machine learning technique for evapotranspiration modeling in various climates. *Atmosphere, 10*, 311.

Van Klompenburg, T., Kassahun, A. & Catal, C. (2020). Crop yield prediction using machine learning: a systematic literature review. *Computers and electronics in agriculture, 177*, 105709.

Walczak, S. (2019). Artificial neural networks. In *Advanced methodologies and technologies in artificial intelligence, computer simulation, and human-computer interaction* (pp. 40-53). IGI global.

Zeynali, M. J., & Hashemi, S. R. (2016). Compare Learning Function in Neural Networks for River Runoff Modeling. *Iranian journal of Ecohydrology, 3*(4), 659-667. doi: 10.22059/ije.2016.60374.