



# Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms

Fatih Tufaner  · Abdurrahman Özbeyaz

Received: 4 May 2020 / Accepted: 2 August 2020 / Published online: 10 August 2020  
© Springer Nature Switzerland AG 2020

**Abstract** Drought, which has become one of the most severe environmental problems worldwide, has serious impacts on ecological, economic, and socially sustainable development. The drought monitoring process is essential in the management of drought risks, and drought index calculation is critical in the tracking of drought. The Palmer Drought Severity Index is one of

the most widely used methods in drought calculation. The drought calculation according to Palmer is a time-consuming process. Such a troublesome can be made easier using advanced machine learning algorithms. Therefore, in this study, the advanced machine learning algorithms (LR, ANN, SVM, and DT) were employed to calculate and estimate the Palmer drought Z-index

## Highlights

- The Palmer Drought Severity Index (PDSI) was modeled to reduce mathematical computational complexity through four machine learning algorithms (LR, ANN, SVM, and DT).
- In the studied models, the meteorological variables were used as input data.
- Palmer's drought computing approach has been re-coded in the Matlab environment. And runoff (RO) and Palmer Index data were obtained by using this software.
- In the study, the best correlation coefficient was obtained in the ANN algorithm with 0.98. The MSE value was 0.40 at this success.
- A novel training data using meteorological variables were developed and shared online.
- By using the developed training data, Palmer drought index values for other regions will be able to be calculated by researchers easily.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10661-020-08539-0>) contains supplementary material, which is available to authorized users.

F. Tufaner  
Department of Environmental Engineering, Engineering Faculty,  
Adiyaman University, Adiyaman, Turkey

F. Tufaner (✉)  
Environmental Management Application and Research Center,  
Adiyaman University, 02040 Adiyaman, Turkey

e-mail: ftufaner@adiyaman.edu.tr

A. Özbeyaz  
Department of Electrical-Electronics Engineering, Engineering  
Faculty, Adiyaman University, Adiyaman, Turkey

values from the meteorological data. Palmer Z-index values, which will be used as training data in the classification process, were obtained through a special-purpose software adopting the classical procedure. This special-purpose software was developed within the scope of the study. According to the classification results, the best  $R$ -value (0.98) was obtained in the ANN method. The correlation coefficient was 0.98, Mean Squared Error was 0.40, and Root Mean Squared Error was 0.56 in this success. Consequently, the findings showed that drought calculation and prediction according to the Palmer Index could be successfully carried out with advanced machine learning algorithms.

**Keywords** Drought · Palmer Drought Severity Index · Regression · Artificial Neural Network · Support vector machine · Linear regression · Decision trees

## Introduction

In recent years, the increase of severity and frequency of the drought and flood disasters, which are considered important extreme climatic events, have caused significant damage to humanity globally. This increasing trend is expected to continue with climate change and to pose more risks to the environment, to the economy and, consequently, to society, depending on water resources (Choubin et al. 2014; IPCC 2014).

A drought is defined as a natural phenomenon that adversely affects land, water resources, and production systems because of precipitation fall significantly below the normal levels, and leading to serious hydrological imbalances (GDWM 2018). It is not easy to determine the onset, duration, and termination of the drought event, as its effects are gradually emerging (Parry et al. 2016). Drought has broad effects on many sub-parts placed under the main headings such as environment, economy, and society. Therefore, the analysis of drought and wet periods with the rainfall-runoff relationship is very important.

Drought monitoring is an important process in dealing with problematic climatic conditions as it enables early warning (Hao et al. 2017). Using the drought indices, which have a complex relationship with climate and environment, is stated as an effective way to detect drought (Alam et al. 2017). It is expressed that 10 indices from a total of 20 indices

are used frequently in the drought analysis of all parts of the hydrological cycle (Wanders et al. 2017). Furthermore, Palmer Drought Severity Index (PDSI) developed by Palmer (1965) and Standardized Precipitation Index (SPI) developed by McKee et al. (1993) are the most commonly used drought indices so far (Tirivarombo et al. 2018). Since drought is dependent on temperature and precipitation, it is stated that PDSI is more suitable for use in assessing the potential impact of climate change on future droughts than SPI (Mishra and Singh 2011). While the SPI is an index computed with only monthly precipitation, the PDSI is calculated based on precipitation and evapotranspiration. In addition, in the calculation of PDSI, how these two first parameters change over time, runoff, moisture supply, and the water holding capacity of the soil at the desired location are used as the input variables (Vicente-Serrano et al. 2010; Wells et al. 2004). PDSI has been widely used and accepted with complaints, criticisms, and improvements until today (Ma et al. 2016). Many studies have modified and improved PDSI (Mo and Chelliah 2006; Yan et al. 2013; Yang et al. 2017; Yu et al. 2019). However, PDSI does not take into account the spatial change of soil, vegetation, topography, and hydrological processes of the basin in drought calculations. It is also calculated by meteorological records on the point scale (Yan et al. 2013). Although the above disadvantages cannot be eliminated, given the aforementioned advantages of the PDSI over the SPI, the PDSI was the drought index chosen to forecast in this study.

With the ongoing global climate change, the development of new and more usable methodologies for assessment of drought conditions and their easier use has become a priority in terms of time constraints and applicability. It is a known fact that SPI is widely used for this reason. However, this raises the issue of reducing the use of indices that have very complex mathematical operations such as PDSI. Although PDSI is modified and updated, it has no widespread solution proposal for the complicated mathematical process (Liu et al. 2017; Mika et al. 2005; Olukayode Oladipo 1985). As drought is important for global climate variability and the need for continuous monitoring, we offer new machine learning model approaches where PDSI drought assessment, which is a versatile drought-monitoring tool, can be used more quickly and easier.

An accurate prediction model trained and evaluated for water scarcity can be an important tool for successful drought and water management (Kisi et al. 2019). During the recent years in which modeling studies have been intensified, different types of drought indexes have been developed and applied with different modeling techniques for drought assessment and monitoring at the regional and global scale. Modeling methods such as Artificial Neural Network (ANN) (Ali et al. 2017; Sigaroodi et al. 2014; Zhang et al. 2019), support vector machine (SVM) (Feng et al. 2019; Roodposhti et al. 2017), linear regression (LR) (Cui et al. 2017; Liu et al. 2019), and decision tree (DT) (Nourani and Molajou 2017; Rhee and Im 2017) algorithms were researched for monitoring, assessing, and forecasting of drought. These data-driven models have become increasingly popular in drought prediction because they are effective in dealing with the non-linear characteristics of PDSI calculation. The PDSI was calculated and estimated from the meteorological data using the advanced machine learning methods in this study, because of Palmer Index has a complex structure with a very long memory. Many drought estimation models using Palmer Index have been developed in literature (Basakin et al. 2019; Mehr and Kahya 2014; Ozger et al. 2012; Rao and Padmanabhan 1984). However, meteorological data were never used as input data in the prediction processes of these studies. Besides, none of these studies presented an approach that calculated the Palmer index value with advanced machine learning methods and meteorological data. Unlike the studies in literature, the Palmer drought index value was predicted and easily calculated using meteorological data in this study. Alternative machine learning methods in PDSI estimation were also applied in the study. Moreover, PDSI computation, which has a complex calculation process, has been made easier with comprehensive and different input data generated from the models. This study presents a pioneering modeling work that will lead to new modification and improvement efforts to address the deficiencies in the PDSI computations.

## Materials and methods

### Case study

Adiyaman province is located (38° 11' and 37° 25' north latitudes and 39° 14' and 37° 31' east

longitudes) within the Middle Euphrates Section of the Southeastern Anatolia Region. The area of Adiyaman province is 7.606 km<sup>2</sup> and its altitude is 669 m. In the north of Adiyaman, there are Malatya Mountains, which are extensions of the Taurus Mountains. The elevation in the region decreases from north to south, and wide flat terrains are seen after the end of the mountains. Atatürk Dam Reservoir, which is 180 km long, 817 km<sup>2</sup> of area and 48.7 km<sup>3</sup> of volume, is located in the southeast of the city. It is Europe's and Turkey's largest dam reservoir and also the sixth largest dam reservoir in the world. The climate of Adiyaman is partly Mediterranean and partly shows continental climate characteristics (DSİ 2014; Tufaner and Dabanlı 2018). Figure 1 shows the location of the study area, its topographic structure, streams and rivers, Atatürk Dam reservoir location, and the location of the meteorological station from which the study data are taken.

The average annual total precipitation amount has been found to be 704 mm based on 56 years (1962–2017) of measured rainfall data at Adiyaman station. During this period, 14 drought incidents occurred, the most severe being in 2017 (368.3 mm). According to SPI 12, 8 moderate, 4 severe, and 2 very severe drought were observed in this period (Tufaner and Dabanlı 2018). This situation leads to negative economic losses and social consequences in this region such as water scarcity, deterioration of water quality, lowering of the Atatürk dam water budget, and agricultural irrigation failures.

### Data

The 1980–2011 meteorological data in Adiyaman province were used in the study. The data were obtained from the Turkish State Meteorological Service (TSMS), (<http://www.mgm.gov.tr/>). The data of monthly average temperature, monthly average actual pressure, monthly average wind speed, monthly average relative humidity, monthly total rainfall, the potential evapotranspiration (PE), available water capacity (AWC), runoff (RO), and PDSI index (Z) were used in the study. In the computation of adjusted potential evapotranspiration, the Thornthwaite formula (Thornthwaite 1948) was

employed. Because it was easily calculated with a small number of inputs (Eqs. 1–4).

$$j_i = \left(\frac{t_i}{5}\right)^{1.514} \quad (1)$$

$$J = \sum_{i=1}^{12} j_i \quad (i = 1, 2, 3, \dots, 12) \quad (2)$$

$$a = (675 \times 10^{-9})J^3 - (771 \times 10^{-7})J^2 + (179 \times 10^{-4})J + 0.492 \quad (3)$$

$$PE_{ad} = 16c \left(\frac{10t_i}{J}\right)^a \quad (4)$$

where  $c$  is a latitude improvement coefficient to account for the different day length between months,  $t_i$  ( $^{\circ}\text{C}$ ) is the average temperature for the  $i$  month,  $j_i$  is the monthly heat index,  $J$  is the annual heat index computed using Eq. 2, and  $a$  is an exponential coefficient calculated as a function of  $J$ .

Available water capacity (AWC) data were obtained from the 1-m soil depth dataset of the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) in the USA (55–109 mm for Adiyaman region) (Webb et al. 2000). In addition, since the AWC value was given as a range of values in the ORNL DAAC, the AWC was calculated according to the following Eqs. (5–6) (Briggs and Shantz 1912) using the average percentage of sand, silt and clay ratios Çelik et al. (2017).

Moisture holding capacity

$$(\text{mm} = (0.03\text{sand}(\%) + 0.35\text{ silt}(\%) + 1.65\text{ clay}(\%)) + 21) \quad (5)$$

Maximum available moisture (mm)

$$= (\text{moisture holding capacity} \times 0.65) + 7 \quad (6)$$

The data used in this study are available to access online and are used to calculate the Palmer drought index in the subsequent studies [data]. These accessible data can be used as training data in machine learning studies related to drought. The data used in the study are summarized in Table 1.

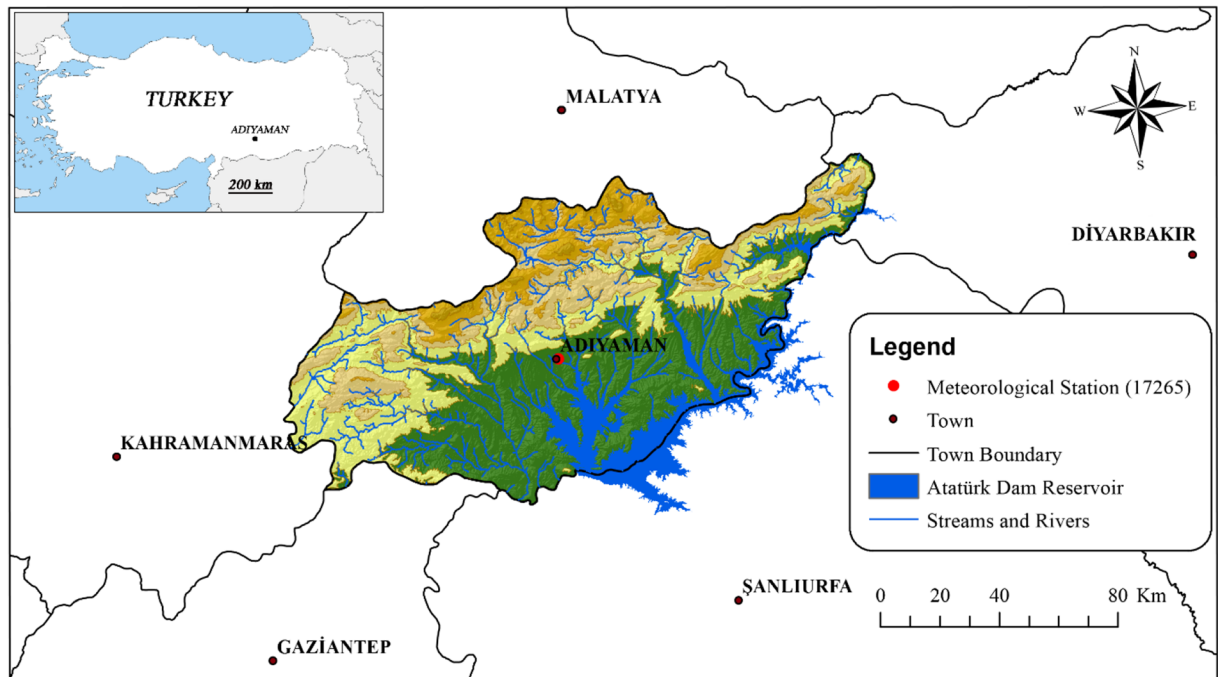
Palmer drought severity index

Palmer Drought Severity Index (PDSI) is a widely used meteorological drought index calculated by soil

moisture and precipitation data of previous months. The positive and negative PDSI values respectively indicate the severity of wet and dry conditions as they move away from 0. Comprehensive analysis taking into account precipitation, surface runoff, and evaporation conditions can be performed with the PDSI calculation based on a long and complex algorithm. And in this way, PDSI is capable of assessing the water potential held in the soil, in other words, drought. The algorithm of PDSI is first developed according to Palmer (1965). In this algorithm, the soil is divided into two parts. The amount of water in the previous month of each section affects the calculation of the following month. Within the scope of the study, special algorithms that can calculate the Palmer drought index by months have been re-developed in the Matlab environment. Special figures and outputs of these algorithms are presented in the following sections. The possible scenarios based on monthly total precipitation and adjusted potential evapotranspiration for the divided parts of the soil are shown in Fig. 2.

PDSI expresses the difference between the observed precipitation amount and the atmospheric evaporating demand (the required precipitation). The required precipitation is calculated using a complex algorithm that takes into account the soil properties. In this algorithm, besides potential evaporation, the duration, the date and the amount of precipitation, and the wetness of the soil determine the direction of the scenario. In the algorithm, potential evaporation is the key forcing factor (Wang et al. 2019).

In Fig. 2,  $P$  and  $PE_{ad}$  represent the total amount of precipitation and evaporation for a month, respectively. According to Palmer, the surface and underlying layers of the soil are both getting wet and dry from the top to bottom when the  $P$  value is either bigger or higher than  $PE_{ad}$  value. Accordingly, if the soil receives more rain ( $5f$ ) than the water holding capacity, more rainfall will flow. This situation is shown as runoff in the scenario (Fig. 2). In the scenario, the water amount entering the soil ( $P < PE_{ad}$ ) is shown by the blue arrows and the water amount exiting from the soil ( $PE_{ad} > P$ ) by the red arrows at the end of the month. Water follows the top to down path while entering and exiting both layers of the soil. For example, in the case of Fig. 2 (7a), if water enters the soil, the water condition in the soil layers will be as in (6c). So, the water wets the top of the top layer. If there is water inlet again next month, (7c) is observed. Also, in case of (7a), if there is water



**Fig. 1** Location of the study area with its topography, streams, and rivers; Atatürk Dam reservoir; and meteorological station

out of the soil, (1f) situation is observed. In other words, the water inlet and outlet at the end of the month are shaped as in the Fig. 2 scenario.

The computation algorithms of the PDSI proposed by Palmer (1965) are developed in the scope of the study. The flowchart of the developed algorithm is shown in Fig. 3.

In Fig. 3,  $T$ : temperature,  $P$ : precipitation,  $PE_{ad}$ : potential evapotranspiration, AWC: available water capacity,  $S_{s-1}$ : available moisture stored in the surface layer at the start of the month,  $S_s$ : available moisture stored in the surface layer at the end of the month,  $S_{u-1}$ : available moisture stored in the underlying layer at the start of the month,  $S_u$ : available moisture stored in the surface layer at the end of the month,  $S$ : available moisture stored in the soil layers at the end of the month, PRO: potential runoff; PR: potential recharge,  $R$ : recharge, PL: potential loss,  $L$ : loss, ET: evapotranspiration, RO: runoff,  $P_{CAFEC}$ : precipitation in climatically appropriate for an existing condition,  $d$ : precipitation excesses and deficiencies,  $D$ : monthly mean of absolute precipitation excess and deficiency values ( $d$ ),  $K$ - $K$ : monthly weighting factors,  $Z$ : the moisture anomaly index,  $U_w$ : effective wetness,  $U_d$ : effective dryness, and PDSI: Palmer Drought Severity Index.

The input data in PDSI is the monthly precipitation sum  $P$ ,  $PE_{ad}$  calculated by Thornthwaite, and AWC

calculated by the soil's sand, clay, and silt ratio. Then, the water potential status is determined for the two layers of soil separated with the input data according to the scenario in Fig. 2. Hydraulic accounting and potential values (PRO, PR,  $R$ , PL,  $L$ , ET, RO) are calculated according to the input data and the water status in the two layers of the soil. In the next step, the CAFEC (Climatically Appropriate For Existing Conditions) coefficients ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) are calculated with these values and  $PE_{ad}$ . In the next step, the monthly precipitation excesses and deficiencies ( $d$ ) are calculated with CAFEC precipitation ( $P_{CAFEC}$ ) and monthly total precipitation ( $P$ ). In addition, the climate characteristic coefficient ( $K$ ) varying according to the measured station is calculated. Therefore, self-calibrating (SC) PDSI which adjusts the coefficients in the PDSI to vary with respect to the measuring station has been developed (Wells et al. 2004). Finally, the moisture anomaly index ( $Z$ ) is calculated and then PDSI is reached.

#### Support vector machine

A support vector machine (SVM) can be used in a regression application by searching an optimum hyperplane between classes. While it is doing that, it maximizes the margin between different groups of data. The



**Table 1** The abstract of the data used in the study

Months (dates)	Monthly average					PE	AWC	Runoff	Z
	Temperature (°C)	Pressure (hPa)	Wind speed (km/h)	Humidity (%)	Total rainfall (mm)	(Inch)	(Inch)	(Inch)	Moisture anomaly index
1 (Jan. 80)	1.9	940.2	21.5	69.4	125.2	0.03	2.33	4.9	-2.07
2 (Feb. 80)	5.2	937.9	13.1	65.8	146	0.2	2.33	5.55	-0.69
3 (Mar. 80)	8.5	935.8	19.7	70.1	186.6	0.63	2.33	0	1.92
10 (Oct. 11)	17.90	938.50	10.80	40.40	40.80	2.41	2.33	0.00	-0.83
11 (Nov. 11)	8.10	940.80	12.70	56.00	90.70	0.47	2.33	3.10	0.48
12 (Dec. 11)	6.50	943.50	9.80	58.80	67.60	0.31	2.33	2.36	-2.01
Avg	17.24	935.92	15.62	50.38	57.77	3.36	2.33	1.47	-0.63
Min	0.40	926.50	5.60	13.80	0.00	0.00	2.33	0.00	-4.17
Max	33.50	944.20	32.20	81.00	348.50	10.53	2.33	13.17	8.66
Std	9.49	4.35	4.03	15.72	64.03	3.13	0.00	2.42	1.71

hyper-plane found by this method is described by the support vectors (Cortes and Vapnik 1995). The first thing to do when applying this method is defining a hyper-plane and maximizing the margin between data groups. In the second step, it extends the hyper-plane defined for the different data groups in non-linearly separable problems. And to do this, it has a penalty term for the misclassification data. In the last step, it maps the data to a high dimensional space where the instances are

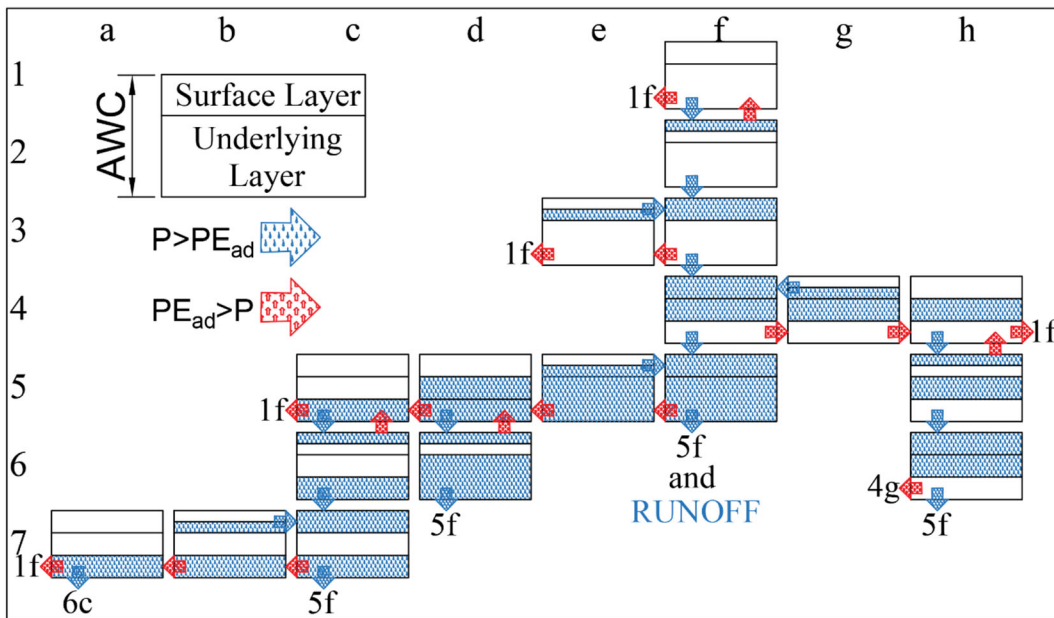
accurately classified. The set of equations that mathematically express this method is given below.

$$wx + b \geq 1, \forall x \text{ of class1 and } wx + b \leq -1, \forall x \text{ of class2}$$

$$\max(2/\|w\|), \text{ and } \min \frac{1}{2} \|w\|^2$$

(7)

In the cases of excellent classification result is not reached, SVM finds the hyper-plane maximizing the



**Fig. 2** Possible scenarios based on monthly total precipitation and adjusted potential evapotranspiration

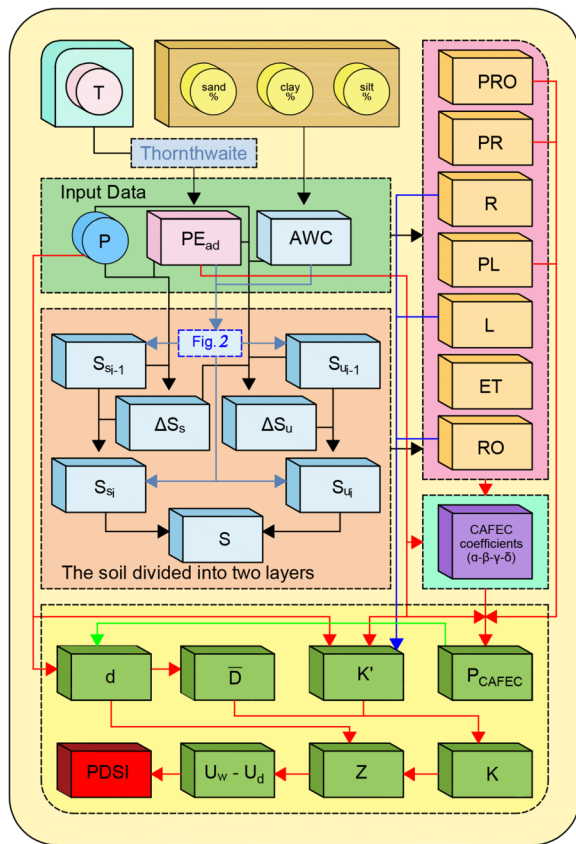


Fig. 3 PDSI computation procedure

margin and decreasing the false-positive rates. To do so, the slack variables ( $\epsilon_i$ ) are kept to zero. While this operation maximize the margin, it does not minimize the misclassification. In that case, SVM uses kernel functions (Meenal and Selvakumar 2018).

$$y_i(wx + b \geq 1 - \epsilon_i), \forall x_i \text{ where } \epsilon_i \geq 0 \text{ and } \min \frac{1}{2} \|w\|^2 + C \sum_i \epsilon_i$$

$$k(x_i \cdot x_j), k \text{ is a kernel function.}$$

(8)

Linear regression

Linear regression is a fundamental method in statistics. In this method, a class of data including features that multiplied by the predefined weights is stated. A linear equation representing such a class is given below.

$$x = w_0 + w_1f_1 + w_2f_2 + \dots + w_nf_n$$

(9)

where  $f_1, f_2, \dots, f_n$  are the feature values in the class  $x$ , and  $w_1, w_2, \dots, w_n$  are the weights. In a regression process, the weight values are estimated from the

training data. A procedure is needed to state the feature values related to each training sample in the process. To give an example, let  $x^k$  is a class for sample  $k$ , and  $f_1^k, f_2^k, \dots, f_n^k$  are the feature values in this class. The weight multiplier of  $f_0$  feature is always one. The estimation equation related to class  $k$  is defined as given below. In the last stage, these estimated values for each training instance are subtracted from actual values.

$$w_0f_0^k + w_1f_1^k + w_2f_2^k + \dots + w_nf_n^k = \sum_{i=0}^j w_if_i^k$$

(10)

Artificial neural network

In order to estimate drought, Multi-Layer-Perceptron (MLP) method (Choubin et al. 2016) known as a feed forward Artificial Neural Network (ANN) model (Sigaroodi et al. 2014) was used in the study. This model works well, when the input is discrete and the output is real. In addition, when this method is used, potential noises on the input are also reduced. The reason why such a method is preferred is that the data is not linearly separated in the classification process. In an MLP neural network, hidden layers are also found except the input and output layers. In such a network, let an input vector is defined as  $X=(x_1, x_2, \dots, x_n)$  and parameter vector is defined as  $W=(w_1, w_2, \dots, w_n)$ ; these vectors are separately multiplied to obtain cross product. The results are summed by bias vector, and they are applied to input of activation function for obtaining the regression output (Li et al. 2019).

There were one input layer, one hidden layer, and one output layer in the ANN model structure. The hidden layer contained four nodes. In the network, sigmoid function was preferred as activation function. MLP method was designed for regression since the data used for the study contains only numeric values. In order to classify the data set, a learning model with backpropagation and supervised was used for the analysis process. In the developed neural network model, the number of validation tests was adjusted to 20. And the learning rate for updating the weights in the model was set to 0.3. The ANN model was concluded with 500 iterations. And finally, 90% and 10% of

the data were randomly selected for training and testing, respectively.

### Decision tree

A decision tree is a hierarchical model for supervised learning in which the local region is defined in fewer steps and recursive branching sequences. A decision tree consists of internal decision nodes and termination branches. This method can be used in studies requiring classification and regression analysis according to the field of application. In regression studies, a tree structure similar to the classification tree is created, but the measure of impurity needed for classification should be replaced with another criterion in regression applications. To express this algorithm mathematically, let it said that  $X_m$  is subset of  $m$ -node to which  $X$  reaches. The  $x \in X$  array satisfies all the conditions in the decision nodes found on the path from the root to the  $m$ -node.

$$b_m(x) = \begin{cases} 1, & \text{if } x \in X_m : x \text{ reaches node } m \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

The branching of a decision tree is decided by the mean squared error obtained from the predicted values. In a regression process, when the  $g_m$  value is taken as the predicted value in  $m$ -node, the obtained equation would be as follows;

$$E_m = \frac{1}{N_m} \sum_{k=0}^n (r^t - g_m)^2 b_m(x^t), \quad N_m = |X_m| \\ = \sum_{k=0}^n b_m(x^t) \quad (12)$$

The  $E_m$  value in the equation is related to the variance in the  $m$ -node. The average of the desired output of the instances reaching a node is used again in this node.

$$g_m = \frac{\sum_t b_m(x^t) r^t}{\sum_t b_m(x^t)} \quad (13)$$

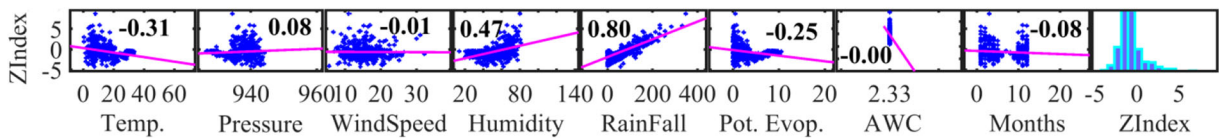
If the error value is acceptable for a node ( $E_m = \theta_r$ ), a leaf node is created and this leaf value stores the  $g_m$  value. Thus, a discontinuous pieced fixed approach model is structured in leaf boundaries. If the error is not of acceptable size, the data reaching to  $m$ -node are further divided so that the sum of the errors in the branches is at the smallest level (Gunaydin et al. 2019).

### Finding and discussions

In this study, the drought estimation models were developed using the advanced machine learning algorithms. In the model, the meteorological data recorded between 1980 and 2011 years belonged to Adiyaman, a province in the southern part of Central Anatolia, Turkey, and the Palmer Z-Index values were used. In the analysis process, four different regression algorithms which were LR, ANN, SVM, and DT were employed. The total number of instances (samples) used in the study was 384. In the regression studies, 85% of these instances were used for training, and 15% for testing. The input variables were monthly average temperature ( $T$ ), monthly average actual pressure ( $P_a$ ), monthly average wind speed (WS), monthly average relative humidity ( $H$ ), monthly total rainfall (RF), potential evapotranspiration ( $PE_{ad}$ ), available water capacity (AWC), and runoff (RO). Besides, Palmer's drought Z-index ( $Z$ ) value was used as the output variable in the model. Palmer drought Z-index values were calculated with the help of a specially developed software in Matlab platform. Detailed information related to the developed software is given in the "Data" section. A linear correlation between inputs and output is shown in Fig. 4.

In Fig. 4, a strong positive correlation is seen between RF and Z-index values, and the  $R$ -value is 0.80. However, while the correlations between  $H$  and Z-index ( $R$ -value is 0.47), and between  $T$  and Z-index ( $R$ -value is 0.31) is weak, the correlations between other inputs and Z-index are weaker. The humidity and temperature are meaningful in the estimation of Z-index, but not for rainfall. In the SVM regression model, the polynomial kernel was used as the kernel function. This algorithm was applied without any hyper-parameter tuning process. When this method was applied, target attributes were normalized as well as the other attributes for determining optimum noise levels easier. Auto-replacing the missing values by global mode variables and auto-converting the nominal instance to numeric ones were applied in the model. Furthermore, kernel caching was turned off, and the number seed was one. The number of decimal places was adjusted as two for the output. The size of the batch was equal to a hundred in the training data set. In the developed model, the training data set was normalized before regression. Before regression, the model was built and regression capability was set to false. In the SVM implementation, the regression analysis was performed by a percentage split method. This process was done as follows: training





**Fig. 4** Pearson correlation matrixes between inputs and output data

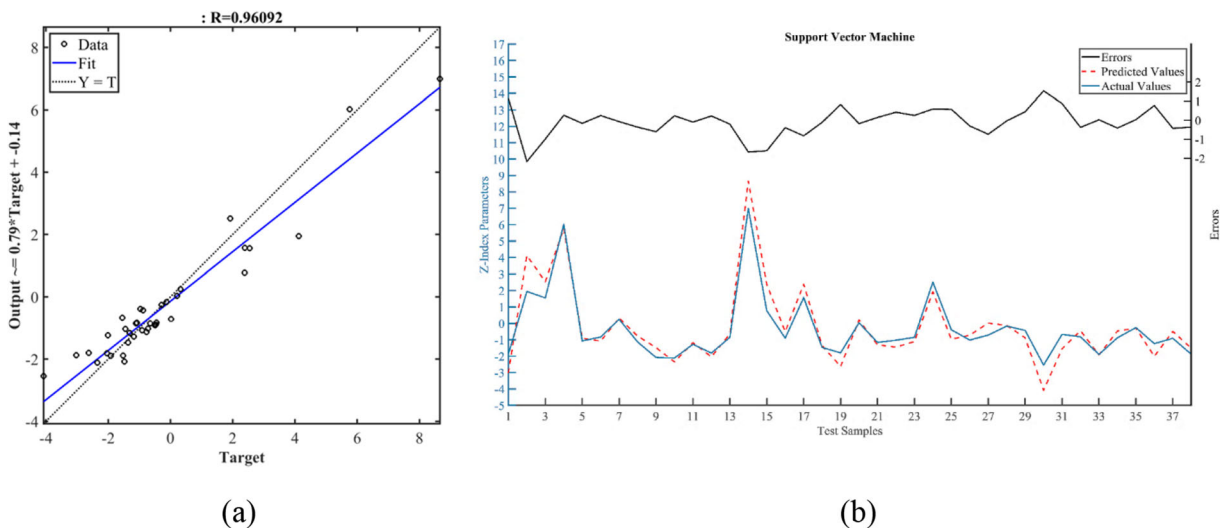
and testing sets were adjusted as 90% and 10% of the all data sets, respectively. It was observed that the SVM was having a robust regression ratio (*R*-value) with 0.92. Besides, in *Z*-index estimation, Pearson’s correlation (*r*) was 0.96; the mean absolute error (MSE) was 0.54, and the root mean squared error (RMSE) was 0.72. Regression plot, prediction values, and error rates are given in Fig. 5a and b, respectively. In Fig. 5b, it is observed that there is a tiny fluctuation in the change between actual and predicted samples. Moreover, the changes in errors are realized at low rates.

In the application of the linear regression method, the Akaike method was used to choose a relative model. In the algorithm, the number of decimal places was set to 4. The preferred number of instances was adjusted to 100 for performing batch prediction in the algorithm. The ridge value was optimized to stabilize degenerate into cases and to reduce overfitting large coefficients. In the model, the attribute selection process was realized by using the M5 decision tree method. In addition, colinear attributes were auto-eliminated in the algorithm. A percentage split approximation was employed to adjust training and testing sets in the method too. 90% of all data were used for training and the rest were used for testing. According to the obtained results, the *R*-value

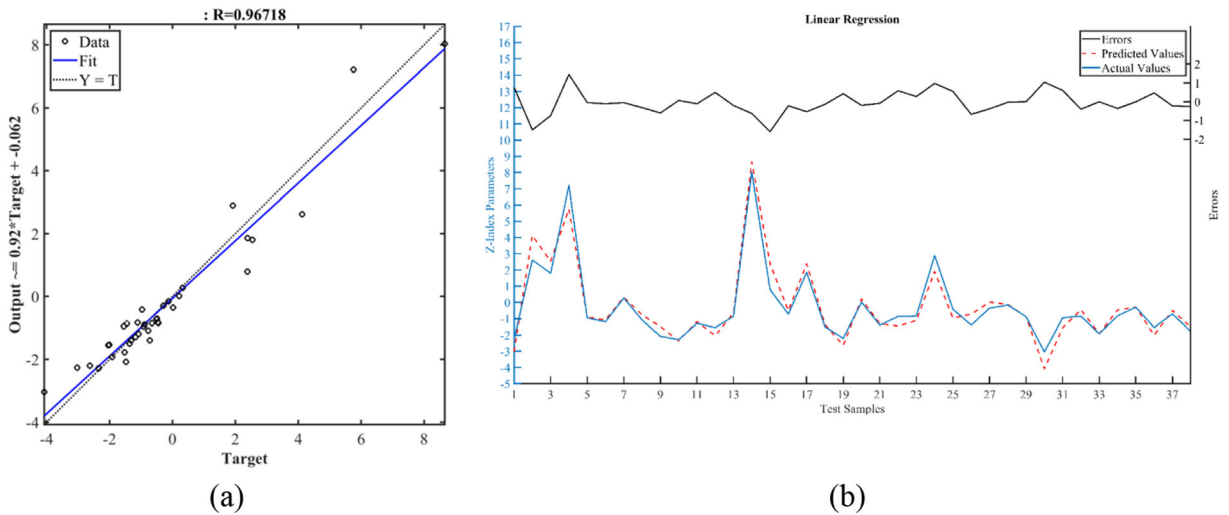
was 0.94; Pearson’s correlation (*r*) was 0.97; MSE was 0.54, and RMSE was 0.72. And the minimum and maximum prediction values varied in low rates. Besides, the worst error rate is 61%, and the best error rate is 100% in *Z*-index estimation. The images summarizing these results are given in Fig. 6. As a result, it can be said that there is a strong correlation between actual outputs and predictions in the LR method. According to this method, the linear estimation model is given in Eq. 14.

$$\begin{aligned}
 Z\text{-Index} = & 0.04xT - 0.11xP + 0.02xWS \\
 & + 0.03xRF + 0.04xPE + 0.04xM \\
 & + 98.67
 \end{aligned} \tag{14}$$

In the ANN regression model, a multi-layer perceptron and the backpropagation algorithm were used to classify instances. The sigmoid equation was the activation function in the model. The initial weights of the connections between nodes were not set in the model. The momentum value of the model was set to 0.2 to update weights. There was one hidden layer in the network. The validation set was adjusted 20 for error could not get worse before training was terminated. Moreover, the normalization process was applied to attributes in the model for improving



**Fig. 5** Regression plot (a), predictions and error values for each test point (b) in SVM

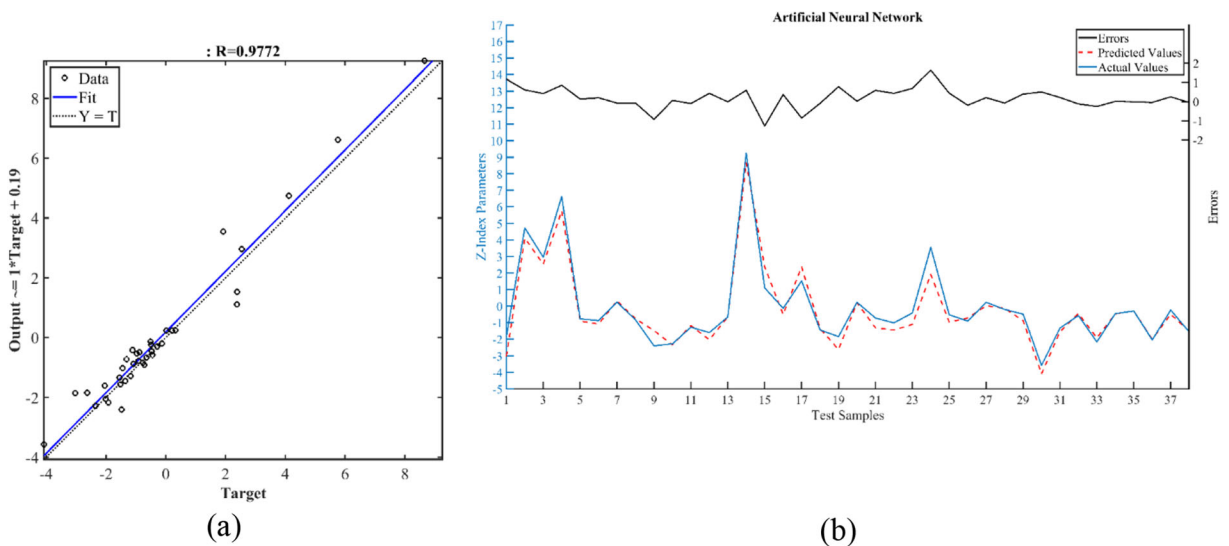


**Fig. 6** Regression plot (a), predictions and error values for each test point (b) in linear regression

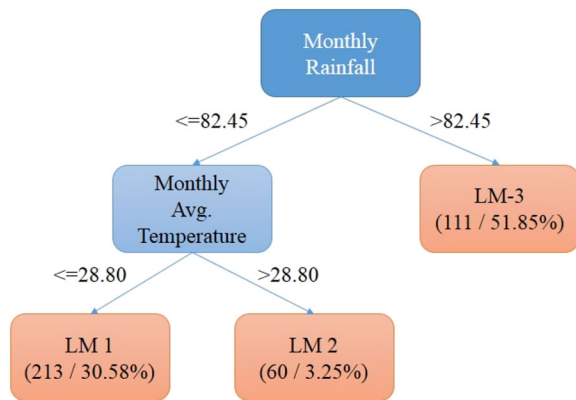
performance in the network. Since instance size was too much, the sample size to be trained at once was chosen as 100. Furthermore, the sample size for validation was zero, and so the network was trained for 500 epochs. Since the instances were too much, the sample size to be trained at once was chosen as 100. The sample size for validation was zero, and so the network was trained through 500 epochs. In the model, the learning rate was 0.3 for updating weight. The regression analysis was performed by a percentage split method. Namely, the rates of training and testing data were adjusted as 90% and 10%, respectively. According to the obtained results, the *R*-value was 0.95, and this rate was the highest among all runs in the present

study. Furthermore, in Z-index estimation, Pearson’s correlation coefficient (*r*) was 0.98; the mean absolute error (MSE) was 0.40, and the root mean squared error (RMSE) was 0.56. The regression plot is shown in Fig. 7a. And the estimation values and error rates are plotted together in Fig. 7b. It is seen from the figures that there is almost no difference between actual and estimated data. Moreover, the error rates are lower in this model than others.

Another regression method used in the study for estimating drought was the Decision Tree algorithm. This algorithm was implemented with a decision-making logic that works from top to bottom. In other words, this algorithm examined the highest level and



**Fig. 7** Regression plot (a), predictions and error values for each test point (b) in ANN



**Fig. 8** Decision tree structure

then the lower levels when dividing the data set used in the study. In this algorithm, the tree was developed to classify the data. The decision tree employed within the scope of the study is shown in Fig. 8.

In the tree structure, the unpruned tree was generated, and the preferred number of instances was 100 to process since batch prediction was being performed. The minimum number of instances was 4 for allowing at a leaf node. The rules (decision list) rather than a tree were generated in the model. These rules are given in Table 2.

The decision-making mechanism of the tree is as follows: if the RF value is higher than 82.45, Rule-3 works. If not and RF value is smaller or equal to 82.45, the algorithm evaluates the *T* parameter. If the *T* value is higher than 28.80, Rule-2 works. If not and *T* value is smaller or equal to 28.80, Rule-1 works. As can be seen from the rules, two parameters come to the fore during the decision-making process. These are RF and *T* values. According to the obtained results, the *R*-value was 0.94, and this rate was the second highest among all runs in the present study. Furthermore, in *Z*-index estimation, Pearson’s correlation coefficient (*r*) was 0.97, the mean absolute error (MSE) was 0.45, and the root mean squared error (RMSE) was 0.61. The regression plot is shown in Fig. 9a. And the estimation values and error rates are plotted together in Fig. 9b. It is seen from the figures that there is almost no difference between

actual and predicted data. Moreover, the error rates are lower in this model than in others.

Drought *Z*-index values obtained according to the Palmer’s drought method were analyzed by using four different regression algorithms. A table comparing the results of these methods is given below (Table 3).

According to these results, Multi-Layer-Perceptron method gave the best performance with 0.95 rate in drought prediction. The correlation coefficient was 0.98 at this success. Besides, it is seen that error values are lower in this success compared with that of other methods. In addition, it is observed that there is no obvious success difference between the methods. When the Palmer index estimation results developed in this study are compared with similar studies in the literature, some original aspects of the application arise in this study. We can compare the originality points of the present study by looking at the table below (Table 4).

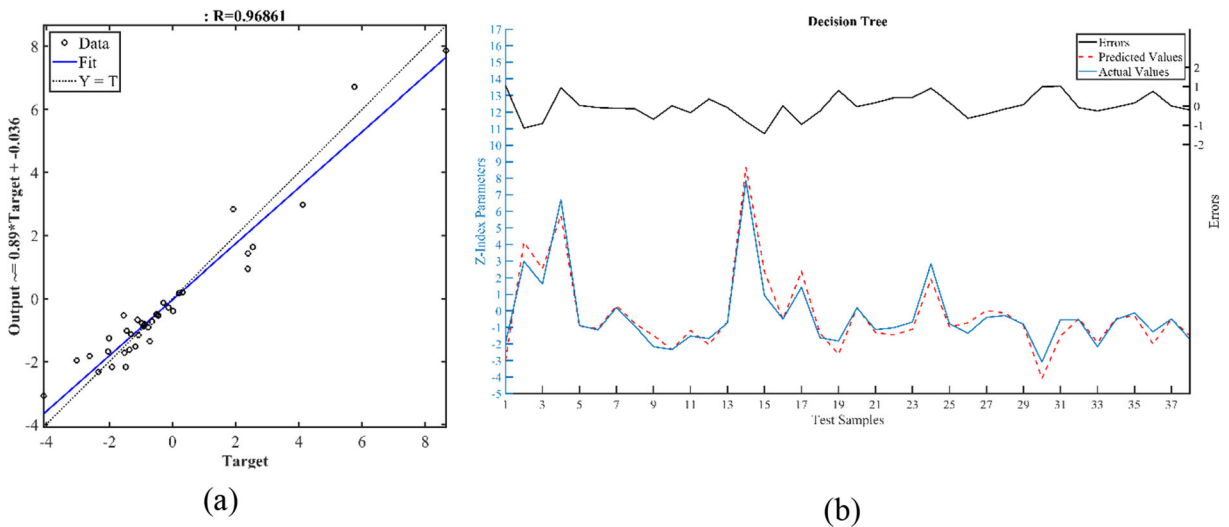
When the table is examined, it is seen that the highest estimation result and the lowest error rate were obtained in the present study. Unlike the literature, meteorological data were used as the input data in the model of the present study. The data developed in our study can be used in the training phases of other drought regression studies, and the Palmer drought index can be easily calculated from the meteorological data for the studied area. Also, the model developed in the study gave a considerable success for drought prediction. However, uncertainties about meteorological data and measurement errors can affect model performance in the study area. For this reason, the data of different stations should be tested in different study areas. In addition to large-scale climate signals, large-scale research is recommended to see if the model forecasts can be improved (Choubin et al. 2014).

**Conclusions**

The estimation of drought with regression modeling is an essential process because some crucial outputs can be

**Table 2** Decision tree rules

$Z - \text{Index} = -0.0022xT - 0.1224 \times Pa + 0.0281xWS + 0.0385 \times RF + 0.1053xPEad + 0.0795xM - 0.2461 \times RO + 110.6418$	Rule-1 (LM-1)
$Z - \text{Index} = -0.0085xT - 0.0305 \times Pa + 0.0043xWS + 0.0093 \times RF + 0.0571xPEad - 0.2929xM - 0.0603 \times RO + 29.5176$	Rule-2 (LM-2)
$Z - \text{Index} = 0.1352xT - 0.0837 \times Pa - 0.0227xWS + 0.0326 \times RF - 0.2452xPEad + 0.0052xM - 0.0139 \times RO + 75.7719$	Rule-3 (LM-3)



**Fig. 9** Regression plot (a), and predictions and error values at each test point (b) in DT

**Table 3** Comparison of the test results according to the estimation methods

	SVM	Linear regression	Multi-layer perceptron	Decision tree
Correlation coefficients (%)	0.96	0.97	0.98	0.97
R-value (%)	0.92	0.94	0.95	0.94
MSE	0.54	0.45	0.40	0.45
RMSE	0.72	0.61	0.56	0.61
Relative absolute error	36.38	29.96	26.67	30.18
Root relative squared error	30.58	25.12	23.02	24.98
Total number of instance in test	38	38	38	38

obtained from the adjusted input values without the need of time-consuming drought calculation with the regression modeling process. In this study, the PDSI parameters were estimated and calculated using four different regression methods, which are frequently used in modeling methodology: LR, ANN, SVM, and DT. These LR and SVM methods were used the first time

in the estimating of PDSI. PDSI is estimated from the input variables: monthly average temperature ( $T$ ), monthly average actual pressure ( $Pa$ ), monthly average wind speed ( $WS$ ), monthly average relative humidity ( $H$ ), monthly total rainfall ( $RF$ ), potential evapotranspiration ( $PEad$ ), available water capacity ( $AWC$ ), runoff ( $RO$ ) and months ( $M$ ).

**Table 4** A comparison of PDSI estimation in the study with its counterparts in the literature

Study	Lead time	Input/output data	Methods	Successes
Present study	Long	Meteorological data/PDSI	ANN, SVM, linear regression, decision tree	0.98 (R-value) 0.40 (MSE value)
Rao and Padmanabhan (1984)	Short	PDSI/PDSI	AR Models	0.92 (MSE value)
Ozger et al. (2012)	Short	PDSI/PDSI	Wavelet and fuzzy logic	0.89 (R-value)
Mehr and Kahya (2014)	Long	PDSI/PDSI	Gene-wavelet model	0.80 (R-value)
Basakin et al. (2019)	Short	PDSI/PDSI	Wavelet-KNN	0.99 (MSE value)

According to the regression results, the best  $R$ -value was obtained with 0.95 when the ANN method was applied. The correlation coefficient was 0.98 at this success. Besides, the MSE value was also the best with 0.40 in this method. Also, in the application of LR and DT methods, the second-best  $R$ -value was obtained with 0.94. And MSE value was 0.45 in these methods. Although the SVM method has the lowest result with 0.92, it can be expressed that this rate is successful in drought prediction studies. As a result, we concluded that PDSI could be estimated at about 90% rates from the meteorological data.

In the study, it was confirmed that the findings obtained from the algorithms were compatible with PDSI results. Besides, the regression methods and input data are shown to facilitate PDSI calculations and will also assist in calculating PDSI parameters for drought analysis in subsequent studies. In this way, input parameters considered to affect drought will be fed into the regression model, and the drought index values will be able to be calculate. In this context, the study offers a pioneering and innovative approach.

## References

- Alam, N. M., Sharma, G. C., Moreira, E., Jana, C., Mishra, P. K., Sharma, N. K., & Mandal, D. (2017). Evaluation of drought using SPEI drought class transitions and log-linear models for different agro-ecological regions of India. *Physics and Chemistry of the Earth*, *100*, 31–43. <https://doi.org/10.1016/j.pce.2017.02.008>.
- Ali, Z., Hussain, I., Faisal, M., Nazir, H. M., Hussain, T., Shad, M. Y., Shoukry, A. M., & Gani, S. H. (2017). Forecasting drought using multilayer perceptron artificial neural network model. *Advances in Meteorology*. <https://doi.org/10.1155/2017/5681308>.
- Basakin, E. E., Ekmekcioglu, O., & Ozger, M. (2019). Drought analysis with machine learning methods. *Pamukkale University Journal of Engineering Sciences*, *25*, 985–991. <https://doi.org/10.5505/pajes.2019.34392>.
- Briggs, L. J., & Shantz, H. (1912). The wilting coefficient and its indirect determination. *Botanical Gazette*, *53*, 20–37.
- Çelik, A., İnan, M., Sakin, E., Büyüç, G., Kırpık, M., & Akça, E. (2017). Changes in soil properties following shifting from rainfed to irrigated agriculture: The Adıyaman case. *Toprak Bilimi ve Bitki Besleme Dergisi*, *5*, 80–86.
- Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., Ahmad, S., & Attarod, P. (2014). Drought forecasting in a semi-arid watershed using climate signals: A neuro-fuzzy modeling approach. *Journal of Mountain Science-England*, *11*, 1593–1605.
- Choubin, B., Malekian, A., & Golshan, M. (2016). Application of several data-driven techniques to predict a standardized precipitation index. *Atmósfera*, *29*, 121–128.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. <https://doi.org/10.1007/Bf00994018>.
- Cui, L. F., Wang, L. C., Lai, Z. P., Tian, Q., Liu, W., & Li, J. (2017). Innovative trend analysis of annual and seasonal air temperature and rainfall in the Yangtze River Basin, China during 1960–2015. *Journal of Atmospheric and Solar - Terrestrial Physics*, *164*, 48–59. <https://doi.org/10.1016/j.jastp.2017.08.001>.
- DSI (2014) General Directory of State Hydraulic Works, Atatürk Dam Reservoir. <http://www.dsi.gov.tr/projeler/ataturk-baraji>. Accessed 18 April 2020
- Feng, P. Y., Wang, B., Liu, D. L., & Yu, Q. (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agricultural Systems*, *173*, 303–316. <https://doi.org/10.1016/j.agsy.2019.03.015>.
- GDWM (2018) Van Lake Basin drought management plan volume-I: General description of the basin and drought analysis vol 1. Flood and Drought Management Department, General Directorate of Water Management (GDWM), T. R. Ministry of Agriculture and Forestry, Ankara.
- Gunaydin, O., Ozbeyaz, A., & Soylemez, M. (2019). Estimating California bearing ratio using decision tree regression analysis using soil index and compaction parameters international. *Journal of Intelligent Systems and Applications in Engineering*, *7*, 30–33.
- Hao, Z. C., Hao, F. H., Singh, V. P., Ouyang, W., & Cheng, H. G. (2017). An integrated package for drought monitoring, prediction and analysis to aid drought modeling and assessment. *Environmental Modelling and Software*, *91*, 199–209. <https://doi.org/10.1016/j.envsoft.2017.02.008>.
- IPCC. (2014). *Climate Change 2014. Impacts, adaptation, and vulnerability*. Cambridge: Cambridge University Press.
- Kisi, O., Choubin, B., Deo, R. C., & Yaseen, Z. M. (2019). Incorporating synoptic-scale climate signals for streamflow modelling over the Mediterranean region using machine learning models. *Hydrological Sciences Journal*, *64*, 1240–1252.
- Li, Y. W., Tang, G. C., Du, J. M., Zhou, N., Zhao, Y., & Wu, T. (2019). Multilayer perceptron method to estimate real-world fuel consumption rate of light duty vehicles. *Ieee Access*, *7*, 63395–63402. <https://doi.org/10.1109/Access.2019.2914378>.
- Liu, Y., Zhu, Y., Ren, L. L., Singh, V. P., Yang, X. L., & Yuan, F. (2017). A multiscalar Palmer drought severity index. *Geophysical Research Letters*, *44*, 6850–6858. <https://doi.org/10.1002/2017GL073871>.
- Liu Q, Zhang S, Zhang H, Bai Y, Zhang J (2019) Monitoring drought using composite drought indices based on remote sensing Sci Total Environ 134585.
- Ma, M. W., Ren, L. L., Singh, V. P., Yuan, F., Chen, L., Yang, X. L., & Liu, Y. (2016). Hydrologic model-based Palmer indices for drought characterization in the Yellow River basin, China. *Stochastic Environmental Research and Risk Assessment*, *30*, 1401–1420. <https://doi.org/10.1007/s00477-015-1136-z>.
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In:



- Proceedings of the 8th Conference on Applied Climatology. vol 22. American Meteorological Society Boston, MA, pp 179–183
- Meenal, R., & Selvakumar, A. I. (2018). Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renewable Energy*, *121*, 324–343. <https://doi.org/10.1016/j.renene.2017.12.005>.
- Mehr, A. D., Kahya, E., & Ozger, M. (2014). A gene-wavelet model for long lead time drought forecasting. *Journal of Hydrology*, *517*, 691–699. <https://doi.org/10.1016/j.jhydrol.2014.06.012>.
- Mika, J., Horvath, S., Makra, L., & Dunkel, Z. (2005). The Palmer Drought Severity Index (PDSI) as an indicator of soil moisture. *Physics and Chemistry of the Earth*, *30*, 223–230. <https://doi.org/10.1016/j.pce.2004.08.036>.
- Mishra, A. K., & Singh, V. P. (2011). Drought modeling - A review. *Journal of Hydrology*, *403*, 157–175. <https://doi.org/10.1016/j.jhydrol.2011.03.049>.
- Mo, K. C., & Chelliah, M. (2006). The modified Palmer drought severity index based on the NCEP North American Regional Reanalysis. *Journal of Applied Meteorology and Climatology*, *45*, 1362–1375. <https://doi.org/10.1175/Jam2402.1>.
- Nourani, V., & Molajou, A. (2017). Application of a hybrid association rules/decision tree model for drought monitoring. *Global and Planetary Change*, *159*, 37–45. <https://doi.org/10.1016/j.gloplacha.2017.10.008>.
- Olukayode Oladipo, E. (1985). A comparative performance analysis of three meteorological drought indices. *Journal of Climatology*, *5*, 655–664.
- Ozger, M., Mishra, A. K., & Singh, V. P. (2012). Long lead time drought forecasting using a wavelet and fuzzy logic combination model: A case study in Texas. *Journal of Hydrometeorology*, *13*, 284–297. <https://doi.org/10.1175/Jhm-D-10-05007.1>.
- Palmer W (1965) Meteorological drought, Research Paper No 45, US Weather Bureau, Washington, DC, 1965.:1-59
- Parry, S., Wilby, R. L., Prudhomme, C., & Wood, P. J. (2016). A systematic assessment of drought termination in the United Kingdom. *Hydrology and Earth System Sciences*, *20*, 4265–4281. <https://doi.org/10.5194/hess-20-4265-2016>.
- Rao, A. R., & Padmanabhan, G. (1984). Analysis and modeling of Palmer's drought index series. *Journal of Hydrology*, *68*, 211–229.
- Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agricultural and Forest Meteorology*, *237*, 105–122. <https://doi.org/10.1016/j.agrformet.2017.02.011>.
- Roodposhti, M. S., Safarrad, T., & Shahabi, H. (2017). Drought sensitivity mapping using two one-class support vector machine algorithms. *Atmospheric Research*, *193*, 73–82. <https://doi.org/10.1016/j.atmosres.2017.04.017>.
- Sigaroodi, S. K., Chen, Q., Ebrahimi, S., Nazari, A., & Choobin, B. (2014). Long-term precipitation forecast for drought relief using atmospheric circulation factors: A study on the Maharloo Basin in Iran. *Hydrology and Earth System Sciences*, *18*, 1995.
- Thomthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review*, *38*, 55–94.
- Tirivarombo, S., Osupile, D., & Eliasson, P. (2018). Drought monitoring and analysis: Standardised Precipitation Evapotranspiration Index (SPEI) and Standardised Precipitation Index (SPI). *Physics and Chemistry of the Earth*, *106*, 1–10. <https://doi.org/10.1016/j.pce.2018.07.001>.
- Tufaner F, Dabanlı İ (2018) Adıyaman İlinde Kuraklık Takibi. Paper presented at the Uluslararası Su ve Çevre Kongresi SUÇEV, Bursa - Türkiye, 22-24 Mart 2018.
- Vicente-Serrano, S. M., Begueria, S., & Lopez-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration. *Index Journal of Climate*, *23*, 1696–1718. <https://doi.org/10.1175/2009JCLI2909.1>.
- Wanders N, Van Loon AF, Van Lanen HA (2017) Frequently used drought indices reflect different drought conditions on global scale. *Hydrol Earth Syst Sci Discuss* 1–16
- Wang, M., Gu, Q. X., Jia, X. J., & Ge, J. W. (2019). An assessment of the impact of Pacific Decadal Oscillation on autumn droughts in North China based on the Palmer drought severity index. *International Journal of Climatology*, *39*, 5338–5350. <https://doi.org/10.1002/joc.6158>.
- Webb RW, Rosenzweig CE, Levine ER (2000) Global soil texture and derived water-holding capacities. Data set. Available online [<http://www.daac.ornl.gov>] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:<https://doi.org/10.3334/ORNLDAAC/548>.
- Wells, N., Goddard, S., & Hayes, M. J. (2004). A self-calibrating Palmer Drought Severity Index. *Journal of Climate*, *17*, 2335–2351. [https://doi.org/10.1175/1520-0442\(2004\)017<2335:Aspsdi>2.0.Co;2](https://doi.org/10.1175/1520-0442(2004)017<2335:Aspsdi>2.0.Co;2).
- Yan, D. H., Shi, X. L., Yang, Z. Y., Li, Y., Zhao, K., & Yuan, Y. (2013). Modified Palmer drought severity index based on distributed hydrological simulation math. *Problems in Engineering*. <https://doi.org/10.1155/2013/327374>.
- Yang, M. Z., Xiao, W. H., Zhao, Y., Li, X. D., Lu, F., Lu, C. Y., & Chen, Y. (2017). Assessing agricultural drought in the anthropocene: A modified Palmer drought severity index. *Water-Sui*, *9*, 9. <https://doi.org/10.3390/W9100725>.
- Yu, H. Q., Zhang, Q., Xu, C. Y., Du, J., Sun, P., & Hu, P. (2019). Modified Palmer Drought Severity Index: Model improvement and application. *Environment International*, *130*, 130. <https://doi.org/10.1016/j.envint.2019.104951>.
- Zhang, R., Chen, Z. Y., Xu, L. J., & Ou, C. Q. (2019). Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province. *China Science of the Total Environment*, *665*, 338–346. <https://doi.org/10.1016/j.scitotenv.2019.01.431>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.