

## A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean

Safwan Mohammed<sup>a,b,\*</sup>, Ahmed Elbeltagi<sup>c</sup>, Bashar Bashir<sup>d</sup>, Karam Alsafadi<sup>e</sup>, Firas Alsilibe<sup>f</sup>, Abdullah Alsالمان<sup>d</sup>, Mojtaba Zeraatpisheh<sup>g,h</sup>, Adrienn Széles<sup>a</sup>, Endre Harsányi<sup>a,b</sup>

<sup>a</sup> Institute of Land Use, Engineering and Precision Farming Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Böszörnyei 138, H-4032 Debrecen, Hungary

<sup>b</sup> Institutes for Agricultural Research and Educational Farm, University of Debrecen, Böszörnyei 138, H-4032 Debrecen, Hungary

<sup>c</sup> Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

<sup>d</sup> Department of Civil Engineering, College of Engineering, King Saud University, P.O.Box 800, Riyadh 11421, Saudi Arabia

<sup>e</sup> School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China

<sup>f</sup> Department of Transport Infrastructure and Water Resources Engineering, Széchenyi István University, 9026 Győr, Hungary

<sup>g</sup> Rubenstein School of Environment and Natural Resources, University of Vermont, 81 Carrigan Drive, Burlington, VT 05405, USA

<sup>h</sup> Gund Institute for Environment, University of Vermont, 210 Colchester Ave, Burlington, VT 05401, USA



### ARTICLE INFO

### ABSTRACT

#### Keywords:

Machine learning  
Data mining  
Drought prediction  
Syria  
Mediterranean basin

Drought is a natural hazard which affects ecosystems in the eastern Mediterranean. However, limited historical data for drought monitoring and forecasting are available in the eastern Mediterranean. Thus, implementing machine learning (ML) algorithms could allow for the prediction of future drought events. In this context, the main goals of this research were to capture agricultural and hydrological drought trends by using the Standardized Precipitation Index (SPI) and to assess the applicability of four ML algorithms (bagging (BG), random subspace (RSS), random tree (RT), and random forest (RF)) in predicting drought events in the eastern Mediterranean based on SPI-3 and SPI-12. The results reveal that hydrological drought (SPI-12, -24) was more severe over the study area, where most stations showed a significant ( $p < 0.05$ ) negative trend. The accuracy of ML algorithms in drought prediction varied in relation to the implementation stage. In the training stage, RT outperformed the other algorithms (Root mean square error (RMSE) = 0.3, Correlation Coefficient ( $r$ ) = 0.97); the performance of the algorithms can be ranked as follows: RT > RF > BG > RSS for both SPI-3 and SPI-12. In the testing stage, both the BG and RF algorithms had the highest correlation  $r$  (observed vs. predicted) (0.58–0.64) and lowest RMSE (0.68–0.88). In contrast, the lowest correlation  $r$  (observed vs. predicted) (0.3–0.41) and highest RMSE (0.94–1.10) was calculated for the RT algorithm. However, BG was more dynamic in drought capturing, with the lowest RMSE and highest correlation. In the validation stage, the BG performance was satisfactory (RMSE = 0.62–0.83,  $r$  = 0.58–0.79). The output of this research will help decision-makers with drought mitigation plans by using the new four machine learning algorithms.

### 1. Introduction

Drought is a natural phenomenon that impacts a wide range of

climatic, hydrological, and ecological systems, all of which have socio-economic implications (Wilhite and Pulwarty, 2017). Drought is primarily caused by meteorological anomalies, in which periods of low

**Abbreviations:** BG, Bagging algorithm; RSS, Random subspace algorithm; R, Random tree algorithm; RF, Random forest algorithm; SPI, Standardized Precipitation Index; SPEI, Standardized Precipitation Evapotranspiration Index; RMSE, Root mean square error; R, Correlation Coefficient; ML, Machine learning algorithms; RVM, Relevance vector machine; ANN, Artificial neural network; KNN, k-nearest neighbors; ELM, Extreme learning machine; SVM, Support vector machine; GP, genetic programming; CoR-SY, Coastal region of Syria; WMO, World Meteorological Organization; M-K test, Mann-Kendall test; SSE, Sen's slope estimator; MAE, Mean Absolute Error; RAE, Relative Absolute Error; RRSE, Root Relative Squared Error.

\* Corresponding author at: Institute of Land Use, Engineering and Precision Farming Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Böszörnyei 138, H-4032 Debrecen, Hungary.

E-mail address: [safwan@agr.unideb.hu](mailto:safwan@agr.unideb.hu) (S. Mohammed).

<https://doi.org/10.1016/j.compag.2022.106925>

Received 7 January 2022; Received in revised form 21 March 2022; Accepted 29 March 2022

Available online 10 April 2022

0168-1699/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

precipitation result in water shortages in various stages of the hydrological cycle or across the entire cycle (McKee et al., 1993). However, drought directly or indirectly affects all ecosystem components (Chaves et al., 2003; Vicente-Serrano et al., 2020). Various factors can influence drought, including a rise in evaporation rates in the atmosphere (Teuling et al., 2013) and atmospheric evaporative demand (Vicente-Serrano et al., 2020), which enhances water shortages and promotes water stress. Therefore, droughts can be understood as a process in which the regional hydrological cycle is pushed to its limit, thereby stressing all other related subsystems (Wilhite et al., 2007). In this context, droughts have affected many parts of the world over the last few decades. For example, in the United States, droughts have affected Texas (2012), California (2012–2015), and the Central Great Plains (2012) (Hao et al., 2018) as well as Australia (1997–2010) (van Dijk et al., 2013), Pakistan (Balochistan drought) (1997–2003) (Ahmed et al., 2016; Durrani et al., 2018), China (Mokhtar et al., 2021; Zhang et al., 2020), Hungary (Harsányi et al., 2021; Alsafadi et al., 2020; Mohammed et al., 2020), Syria (Mohammed et al., 2020; Mohammed et al., 2021), and the Mediterranean (Kreft et al., 2007).

Drought prediction is one of the most challenging issues for climate scientists and hydrologists due to its complexity and development on a spatial-temporal scale (Hao et al., 2018). Drought events are usually predicted using statistical, dynamical (Mariotti et al., 2013; Mishra and Singh, 2011; Shahid, 2010), and hybrid models (Pozzi et al., 2013). In statistical models, correlation relationships between climate variables and drought indicators are used for drought prediction (Yaseen et al., 2015). Unlike statistical models, dynamical models are based on the physical interconnections between earth, ocean, and climate. To develop drought forecasts and predictions, these interactions are theoretically defined and resolved in dynamical models (Turco et al., 2017). In contrast, a hybrid model is a combination of statistical and dynamical methods (Murakami et al., 2016; Strazzo et al., 2019). For example, to develop an ensemble prediction, numerous dynamical model predictions can be integrated using a statistical framework, which gives weight to the distinct dynamical model predictions (Madadgar et al., 2016). However, statistical models are widely employed to predict droughts due to their simplicity (Belayneh et al., 2016) and inexpensive processing requirements (Ganguli and Reddy, 2014; Mariotti et al.,

2013; Xu et al., 2018). Machine learning (ML) techniques have also recently been used to predict drought in several parts of the world.

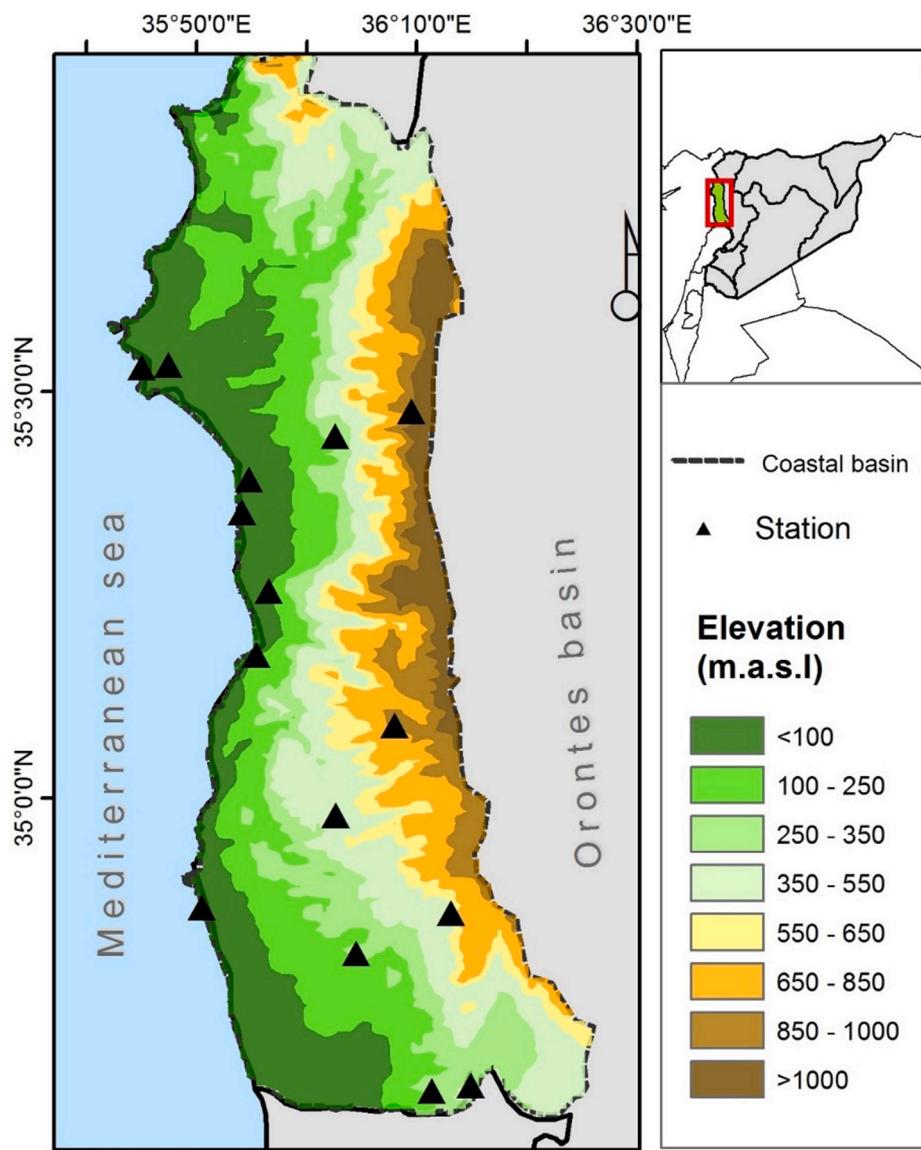
Machine learning (ML) approaches involve a set of commands that allow systems to learn and improve from historical data without requiring extensive programming (Sachindra and Kanae, 2019). In various climatological applications, such as rainfall prediction, various machine learning algorithms have been used to develop models that can reproduce the empirical relationship between the different variables (Parmar et al., 2017), drought prediction (Tian et al., 2018), forecasting heat waves (Khan et al., 2019), and runoff simulation (Yaseen et al., 2015). Some of the most extensively used ML algorithms for modeling the relationship between different variables include the relevance vector machine (RVM), the artificial neural network (ANN), the k-nearest neighbors (KNN), the extreme learning machine (ELM), the support vector machine (SVM), genetic programming (GP), and random forests (RF) (Yaseen et al., 2015; Fahimi et al., 2017; Deo and Şahin, 2015; Bourdin et al., 2012; Rhee and Im, 2017; Nourani et al., 2014; Wang et al., 2009). In this context, many ML techniques have been developed to model the complex non-linear interactions among drought factors and predictors such as SVM, regression trees, ANN, ELM, and RF (Ganguli and Reddy, 2014; Granata, 2019; Morid et al., 2007; Mishra and Desai, 2006; Mouatadid et al., 2018; Feng et al., 2019). Additional examples of the use of ML algorithms in drought prediction can be found in Table 1.

The Mediterranean region was highlighted as being a climate change hotspot, with winter rainfall dropping by up to 40% (Ginoux et al., 2004; Tuel and Eltahir, 2020). Drought episodes in the Mediterranean, particularly those that occur during the wet season, can significantly influence water supplies through a decrease in groundwater levels and the available water in dam reservoirs and lakes (Lorenzo-Lacruz et al., 2013, 2017; Raymond et al., 2016). In this context, water shortage can have a negative impact on numerous economic sectors, wild biodiversity, and agricultural productivity, especially in countries that significantly rely on rain-fed agriculture (Turkes et al., 2020; Schilling et al., 2020).

Syria, which is situated in the Mediterranean region, is subjected to climate change, especially drought (Mohammed et al., 2020; Mohammed et al., 2021). Previously, Mohammed et al. (2020) reported an increase in drought trend across the coastal region of Syria (CoR-SY) (i.e.,

**Table 1**  
An overview of Machine learning methods for drought prediction in some parts of the world.

Machine learning algorithms	Study area	Drought	Performance	Reference
Extreme learning machine (ELM) and artificial neural network (ANN)	Australia	drought prediction	ELM was better than ANN	(Deo & Şahin, 2015)
recursive multistep neural network (RMSNN), direct multistep neural network (DMSNN)	Australia	Drought Forecasting	The RMSNN better 2–3 months forecasting, DMSNN better than RMSNN for forecast lead times of 4–6 months	(Barua et al., 2012)
least squares support vector machine (LSSVM) M5Tree	Australia	modelling of the SPI	M5Tree model was better than the LLSVM model	(Deo et al., 2017)
wavelet neural network (WNN) the autoregressive integrated moving average (AIRMA)	Australia	Drought prediction	AIRMA model showed an obvious advantage over the SVM and WNN models	(Deo et al., 2018)
support vector machines (SVMs)				
multilayer perceptron neural network (MLPNN)	Pakistan	drought forecasting	MLPNN has potential capability for SPEI drought forecasting	(Ali et al., 2017)
autoregressive integrated moving average (ARIMA)	India	drought indices predication	LSTM better than ARIMA	(Poornima and Pushpalatha, 2019)
Long short-term memory (LSTM)				
artificial neural network (ANN)	Iran	drought forecasting	SVM the best performance	(Mokhtazar et al., 2017)
Adaptive neuro-fuzzy interface system (ANFIS)				
support vector machine (SVM)				
Support Vector Machine (SVM)	India	prediction of effective drought index (EDI)	SVR-HHO better than SVR-PSO	(Malik et al., 2021a)
Particle Swarm Optimization (PSO)				
Harris Hawks Optimization (HHO)				
support vector regression (SVR)	India	meteorological drought prediction	SVR-GWO outperformed SVR-SHO	(Malik et al., 2021b)
Grey Wolf Optimizer (GWO)				
Spotted Hyena Optimizer (SHO)				



**Fig. 1.** Study area and selected meteorological stations.

the eastern Mediterranean), where the CoR-SY was represented by 4 weather stations (1990–2010). Similarly, Mathbout et al. (2018) assess drought variability in Syria using the SPI and the Standardized Precipitation Evapotranspiration Index (SPEI), nonetheless, the CoR-SY was covered by 4 weather stations from 1961 to 2012. However, there is still a lack of literature about drought evolving in the eastern Mediterranean. On the other hand, the current conflict in Syria has led to a malfunction in many climate stations across the country. The ML algorithms could be one of the solutions to overcome the gap in climate data, based on the historically recorded data. However, to the best of the authors' knowledge, the performance of ML algorithms was never applied in Syria for drought prediction and assessment. Thus, this research aimed to bridge the gap in the literature about the drought evolution in the CoR-SY and to evaluate the performance of ML algorithms in drought prediction. **The detailed goals were to i) assess drought evolution in the CoR-SY based on observed rainfall retrieved from 15 stations covering the whole CoR-SY, and ii) to assess the applicability of the bagging (BG), random subspace (RSS), random tree (RT), and random forest (RF) in drought prediction across the CoR-SY.**

## 2. Materials and methods

### 2.1. Study area and data collection

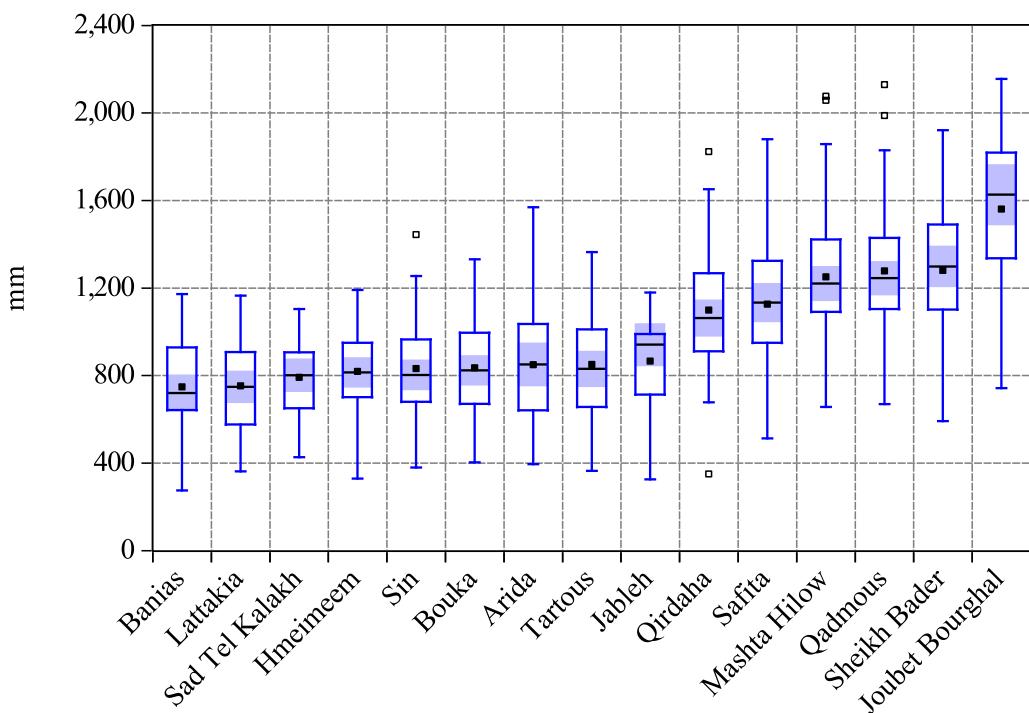
The study area is located in the eastern part of the Mediterranean basin (west of Syria) between  $35^{\circ}.49'$  to  $36^{\circ}.31'$  E and  $34^{\circ}.49'$  to  $36^{\circ}.05'$  N (Fig. 1), covering an area of  $5274 \text{ km}^2$ . The climate is a typical Mediterranean one (Csa and Csb), with a cold and rainy winter and a hot and dry summer. The average annual temperature ranges between  $14.8^{\circ}\text{C}$  (mountains) and  $19.3^{\circ}\text{C}$  (plain), while the average rainfall ranges from 750 mm (plain) to 1250 mm (mountains). In terms of topography, the study area can be divided into three geomorphological groups: the mountain area (400–1700 m), which dominates the eastern part; the hillslopes (100–400 m); and the plain (0–100 m), which is located near the coast. However, inclination ranges between 0 and  $-60^{\circ}$ . The study area consists of two governorates, Tartous and Lattakia, where the total population is 3 million.

Available rainfall data about the CoR-SY was collected from the meteorological department in Syria. The stations were not chronologically consistent with the first station established in 1946, while new stations were established after the 1970 s. Additionally, some stations were in service for a period of time and then went out of service due to

**Table 2**

some characteristics of monitoring climate stations in the study area.

Station	Abbreviation	Lat	Long	year	Monitoring period	Elevation (m)
Bouka	BO	35° 32'	35° 48'	1946–2005	60	50
Lattakia	LAT	35° 32'	35° 46'	1952–2005	55	7
Joubet Bourghal	JB	35° 30'	36° 11'	1960–1990	30	950
Qirdaha	QI	35° 27'	36° 04'	1959–2005	47	300
Hmeimeem	HM	35° 24'	35° 56'	1956–1991	35	48
Jableh	JA	35° 22'	35° 55'	1976–1996	21	14
Sin	SN	35° 15'	35° 58'	1957–2001	45	40
Banias	BA	35° 13'	35° 57'	1974–2004	31	5
Qadmous	QA	35° 06'	36° 09'	1959–2004	46	750
Sheikh Bader	AB	34° 59'	36° 05'	1960–2004	45	550
Mashta Hilow	MH	34° 53'	36° 16'	1959–2004	46	500
Tartous	TR	34° 53'	35° 52'	1957–2005	49	5
Safita	AF	34° 49'	36° 08'	1959–2005	47	350
Sad Tel Kalakh	AT	34° 41'	36° 16'	1972–2001	30	225
Arida	AR	34° 40'	36° 19'	1958–1998	41	240

**Fig. 2.** Boxplot of yearly rainfall in the selected meteorological stations (mean (■), median (—), near outliers (□), median 95% confidence (shade)).

technical or other issues. Nevertheless, data were collected from 15 stations, the lowest of which had a monitoring period of 21 years, which is appropriate for drought monitoring and ML application (Table 2, Fig. 2).

## 2.2. SPI and drought trends

In this study, the SPI (McKee et al., 1993) was used for drought calculation, which was recommended by the World Meteorological Organization (WMO) (Mohammed et al., 2020). The SPI is a flexible index that can give an overview of drought events on different timescales, such as the SPI-1 and SPI-3 for agricultural drought and the SPI-6, SPI-9, and SPI-12 for hydrological drought. Statistically, the SPI structure transforms rainfall values to one of the probability distributions such as gamma, Pearson III, or log-logistic. Data are then converted to a normal distribution (mean = 0, variance = 1). The SPI calculation was extensively reported by many authors in the literature (i.e., Karavitis et al., 2011; Angelidis et al., 2012); thus, we limited this section to the essence of the concept of the SPI. The SPI can be interpreted as follows:

$$SPI = \frac{R_i - \bar{R}}{\sigma}, \quad (1)$$

where  $R_i$ : rainfall (mm) in a year (i),  $\bar{R}$ : rainfall average, and  $\sigma$ : standard deviation.

In this context, the wet period represents the positive value of the SPI, while the negative value represents the water shortage (Jasim and Awchi, 2020). The SPI categories are shown in Table 3.

## 2.3. Trend analysis

In environmental research, many methods can be implemented to track the evolution of trends during any given time period, including parametric and non-parametric methods (Gioia et al., 2020). In this research, the Mann–Kendall test (M–K test) (Mann, 1945; Kendall, 1975) and Sen's slope estimator (SSE) (Sen, 1968) was adopted for tracking drought trends in the studied climate stations. The M–K is used as one of the non-parametric trend tests, which means that the data does not have to follow a certain statistical distribution. The M–K test was

**Table 3**

SPI categories for drought classification.

Drought group	SPI values
Normal	0 <
Mild drought	0 to -0.5
Moderate drought	-0.5 to -0.84
Severe drought	-0.84 to -1.28
Extreme drought	-1.28 to -1.65
Very extreme drought	< -1.65

chosen for this research because it can detect any trend, whether it is linear or non-linear (Wu et al. 2008). However, both M-K, and SSE are widely used for capturing drought trends and magnitude.

For the time series studied ( $x_1, \dots, x_n$ ), the M-K S statistics (Sharma and Goyal, 2020) are as follows:

$$M - KS = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sig}(x_j - x_i) \quad (1)$$

where  $x_i$  and  $x_j$  were the data collected in years j and k ( $j > k$ ), n covered the whole period, and  $\text{sig}(x_j - x_i)$  is as follows:

$$\text{sign}(x_j - x_i) = \begin{cases} 1 & x_i < x_j \\ 0 & x_i = x_j \\ -1 & x_i > x_j \end{cases} \quad (2)$$

When  $n \geq 8$ , we have an almost normal distribution for the M-K S statistics; thus, we can use the following hypothesis:

$$E(M - KS) = 0 \quad (3)$$

$$VAR(M - KS) = \frac{n(n-1)(2n+5) - \sum_{z=1}^y t_z(t_z-1)(2t_z+5)}{18} \quad (4)$$

In this case, y represents the tied groups, and  $t_z$  is its size. Finally, the statistic of MK is as follows:

$$MK \left( \omega \right) = \begin{cases} \frac{VAR(M - KS) - 1}{\sqrt{VAR(M - KS)}} & VAR(M - KS) > 0 \\ 0 & VAR(M - KS) = 0 \\ \frac{VAR(M - KS) + 1}{\sqrt{VAR(M - KS)}} & VAR(M - KS) < 0 \end{cases} \quad (5)$$

The positive or negative trend of the M-K S statistics indicates the increase or decrease in the value of the observed variable, where the H0 of MK ( $\omega$ ) was tested at a 0.05 confidence level.

On the other hand, the magnitude of a trend within the time series studied was determined using Sen's slope estimator (SSE) (Sen, 1968). The SSE is a non-parametric method that can be understood as follows:

$$SSE = \frac{\Gamma_j - \Gamma_k}{j - k} \quad \text{for all } j < k \quad (6)$$

where  $\Gamma_j - \Gamma_k$  are the values of data (for more information, see Sen, 1968).

In this study, we adopted the M-K and SSE statistical analysis methods for detecting changes in a time series and to highlight whether these changes are significant. Finally, the results were visualized using the EViews program and the Geographical Information System (GIS).

#### 2.4. Machine learning methods for predicting drought events

Drought conditions were predicted using four rule-based machine learning approaches: the bagging algorithm (BG), random subspace (RSS), random forest (RF), and random tree algorithms (RT), as described in Table 4. All the modeling work was performed in WEKA (v.3.8.4) Software. The WEKA Software is a set of data mining-related ML techniques. It can handle and process different data types through different tools (i.e., regression, clustering, and visualization). The

**Table 4**

The parameters of the machine learning algorithm used for SPI 3 and SPI 12 modeling.

Model name	Description of parameters
Random Tree (RT)	Batch size-100, seed = 1, minimum variance proportion = 0.001
Bagging (BG)	Batch size-100, bag Size percent = 100, Classifier = REPTree, max depth = 0, numbers of executions slots = 1, number of iteration = 10, random seed = 1
Random Forest (RF)	Batch size-100, bag Size percent = 100, max depth = 0, numbers of executions slots = 1, number of iteration = 100, random seed = 1
Random Subspace (RSS)	Batch size-100, Classifier = REPTree, random seed-1, subspace size = 0.5, numbers of executions slots = 1, number of iteration = 10

program also includes a graphical user interface (GUI), which is one of the advantages that help users in running the program. There are other alternatives such as MATLAB, Python, and R but they need much time to prepare and implement computer codes. There is no effect on the results obtained. We selected the easy process for modeling compared to other software. In this sense, some parameters, especially classifiers such as M5P, Additive regression (AR), Gaussian process regression (GPR), REPTree, and SMO-Support Vector Machine (SMO-SVM), were initially implemented. These parameters consider as the major parameter for getting the modeling results. The output classified REPTree as the best classifiers compared to other alternatives.

##### 2.4.1. Bagging algorithm (BG)

Breiman (1996) introduced the bagging algorithm, which used a bootstrap sampling technique. Bagging is a straightforward and effective ensemble approach that produces new  $m$  training sets and then fits the datasets using  $m$  models. Their prediction results are combined by averaging the output or voting. Boosting is a term that refers to a collection of algorithms that use weighted averages to transform underperforming learners into stronger learners (Breiman, 1996). In this work, regression was primarily performed, while bagging was applied to a variety of sophisticated model architectures, regression trees, and variable selection. Various regression trees were combined to generate a single output with this bagging technique using a weighted average. Multiple exercise datasets of comparable size were randomly selected from a domain during the model's development stage. A regression tree model was also created for each of the input datasets. Each tree was unique and produced a distinct forecast based on the changes in the training dataset. Finally, the weighted average of each regression tree's projections was measured. The settings chosen in this procedure are described in Table 4.

##### 2.4.2. Random subSpace (RSS)

Ho (1998) created the RSS model as a novel ensemble machine learning approach for tackling environmental concerns. This model's multiple classifiers are integrated and trained on a different feature space. This approach creates numerous training subsets for the classifiers, which serve as the training foundation. Unlike other ensemble models, the RSS model uses many samples on function space rather than example space, as indicated by Havlek et al. (2019). This approach utilizes bootstrapping and grouping. When various algorithms such as artificial neural networks, decision trees, or similar approaches are utilized, random subspace models may show non-linear correlations and probable interactions between features (Zhao and Liu, 2007). In the random subspace method, the classifiers are created using the random subspace approach of the data feature space. These classifiers are often mixed by a simple majority vote when it comes to the final decision. High computational costs are a significant drawback of this approach (He et al., 2003). The parameters used for implementing this approach are presented in Table 4.

**Table 5**

Details of indices with mathematical formulations for validation of data mining techniques.

Indices	Formula*	Reference	Range	Ideal level	Description
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (SPI_A^i - SPI_P^i)^2}$	Ulgen, & Hepbasli (2002)	[0 to $\infty$ ]	0	Depicts the differences between observed values and estimated one.
Mean Absolute Error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^N  SPI_P^i - SPI_A^i $	Vishwakarma et al. (2022)	[0 to $\infty$ ]	0	Evaluates the average size of errors (I.e., $SPI_P^i$ vs. $SPI_A^i$ )
Relative Absolute Error (RAE)	$RAE = \left  \frac{SPI_A^i - SPI_P^i}{\overline{SPI}_P} \right  \times 100$		[0 to $\infty$ ]	0	Evaluates the ML performance
Root Relative Squared Error (RRSE)	$RRSE = \frac{\sqrt{\sum_{i=1}^N (SPI_A^i - SPI_P^i)^2}}{\sqrt{\sum_{i=1}^N (SPI_A^i - \overline{SPI})^2}}$	Khan et al. (2020)	[0 to $\infty$ ]	0	Calculates the total squared error and normalizes it by dividing it by the total squared error.
Correlation Coefficient (r)	$r = \frac{\sum_{i=1}^n \{(SPI_A^i - \overline{SPI}_A)(SPI_P^i - \overline{SPI}_P)\}}{\sqrt{\sum_{i=1}^n (SPI_A^i - \overline{SPI}_A)^2} \sqrt{\sum_{i=1}^n (SPI_P^i - \overline{SPI}_P)^2}}$	Pearson (1896)	[-1 to +1]	+1	Shows the similarity between observed and expected values

\* $SPI_A^i$ : actual value,  $SPI_P^i$ : predicted value,  $\overline{SPI}$ : the mean value of reference samples, and N is the total number of data points.

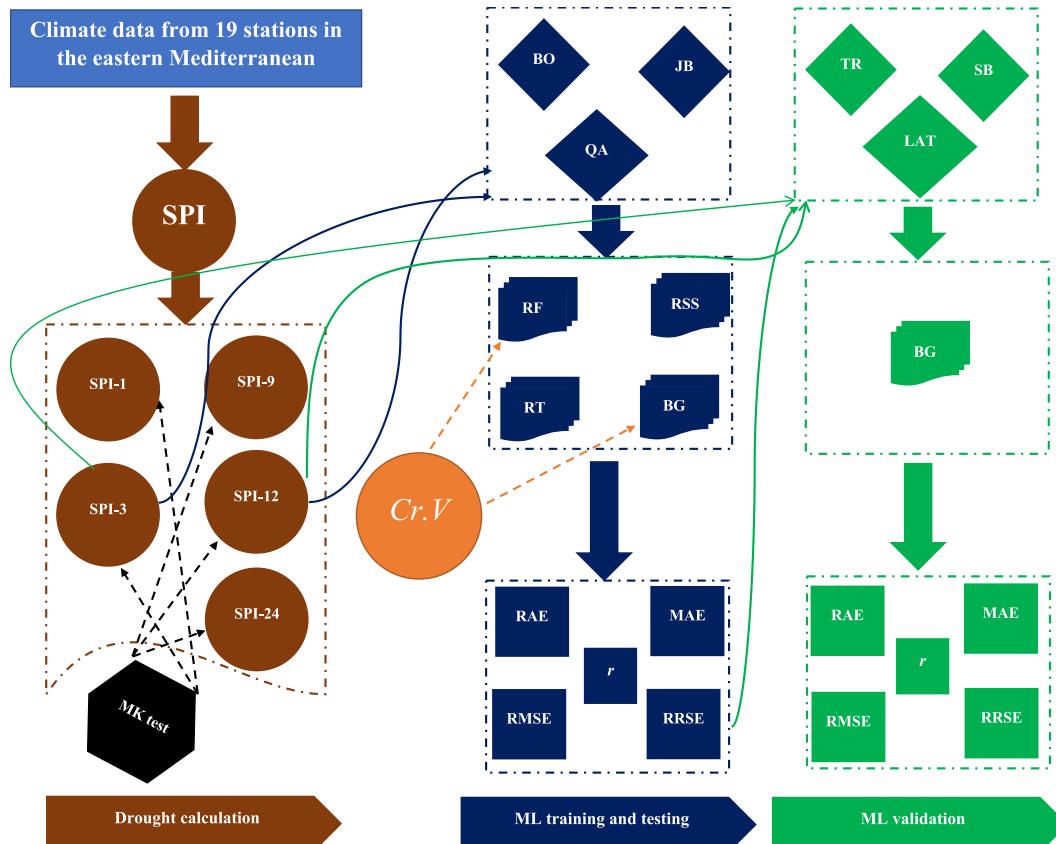
#### 2.4.3. Random forest (RF)

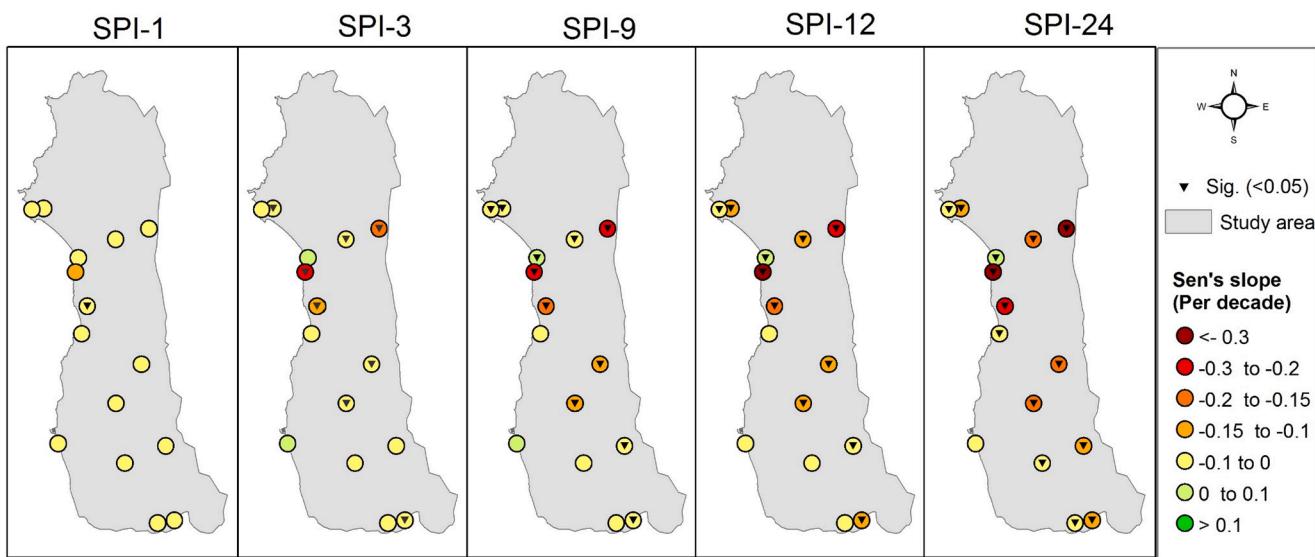
The random forest (RF) model is an ensemble method based on the well-known classification and regression tree method (CART) (Breiman et al., 1984). The ensemble is created by averaging many trees from the training data using various bootstraps drawn from the training data. Furthermore, only a random selection of variables is assessed at each node. RF with a high number of trees is resistant to overfitting and noise, is non-informative, and has correlated features. The RF technique consists of two phases. In the first phase, each tree in the random forest (consisting of several decision trees, e.g., 1000) is generated independently using bootstrap sampling from the original data. Each tree is sampled using a separate bootstrap sample, which is about 66 percent of

the total data, n. Second, those observations (33%) that are not included in the bootstrap sample are referred to as out-of-bag, and the percentage of misclassified samples (known as out-of-bag error) is a measure of the method's accuracy (Peters et al., 2007). Over the last decade, the RF model has been used in various environmental studies and other environmental issues (Peters et al., 2007; Wiesmeier et al., 2011; Taghizadeh-Mehrjardi et al., 2016). The RF approach makes no assumptions about the model error distribution or any special relationship within the drought index series.

#### 2.4.4. Random tree (RT)

The RT algorithm was established using a conventional method and

**Fig. 3.** Flowchart for different steps adopted in this research.



**Fig. 4.** Drought trend over the study area for different time scales (-1,-3,-9,-12,-24 months) by using SPI.

then refined using decision trees on an arbitrary column subdivision. This model was created using a single significant exclusion approach similar to traditional trees. The model is very fast and adaptable for a tree starter and may be used to illustrate a wide range of issues (Najock et al., 1982; Busch et al., 2009). A supervised classifier is a collaborative learning model that is used to construct more distinct origins. It provides a bagging impression to conceptualize an arbitrary information standard for developing decision tree models. Each node in a typical tree is separated using the smallest fragment in the entire variable. All the nodes are split by the smallest fragment between a subset of forecasters that were arbitrarily chosen at a node. In a decision-making tree, each leaf receives an optimal linear model to a local subspace, with each leaf bearing an explanation. The use of decision trees alone has substantially advanced random forests; tree variation is formed when two trees go in different directions (Amit and Geman, 1997). The parameters chosen for this model are described in Table 4.

#### 2.4.5. Model performance

The ML algorithm's performance was evaluated by comparing the estimated values of the SPI (i.e., modeled values from through ML algorithms) with the observed values in different time scales. The model's performance was evaluated using cross-validation (*Cr.V*) (Bouras et al., 2021). The *Cr.V* is one of the simple techniques used for selecting and evaluating algorithms, which avoid over-fitting (Bouras et al., 2021). However, the *Cr.V* at 5 folds for the best models was carried out.

In literature, there is a wide range of statistical indicators that could be used to evaluate the performance of the ML algorithm. However, in this research, we adopted the most common statistical indicators (Root mean square error (RMSE), Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE), Correlation Coefficient (*r*)), along with the Taylor diagram, as depicted in Table 5. The adopted statistical indicators depict the differences between observed values and estimated one along with the average size of errors.

In this study, three stations with different elevations (BO (50 m), JB (950 m), and QA (750 m)) were chosen for model training (70% of the dataset) and testing (30% of the dataset). The performances of the BG, RSS, RF, and RT algorithms were assessed by using the indicators in Table 5. The algorithms that achieved the best performance were then implemented in another three stations with different elevations (LAT (7 m), SB (550 m), and TR (5 m)) for validation, as presented in Fig. 3.

## 3. Results

### 3.1. Drought trends over the study area

For agricultural drought (i.e., SPI-1 and SPI-3), few stations depicted a significant ( $p < 0.05$ ) negative trend of drought across the study area. However, the majority of the stations experienced negative but non-significant drought (Fig. 4). For SPI-1, 60% of the stations depicted a negative trend, while 40% showed no trend (Sen's slope = 0). In the case of SPI-3, 40% of the stations depicted a significant negative ( $p < 0.05$ ) trend, while 34% showed a negative but not significant drought trend (Fig. 4).

In terms of hydrological drought (SPI-9, -12, and -24), the majority of the stations showed a significant negative ( $p < 0.05$ ) trend (Fig. 4). In this regard, 54%, 74%, and 74% of the total stations showed a significant negative ( $p < 0.05$ ) trend for SPI-9, SPI-12, and SPI-24, respectively. Interestingly, the drought magnitude was more pronounced in the eastern (far away from the coast) and northern parts (near the Turkish border).

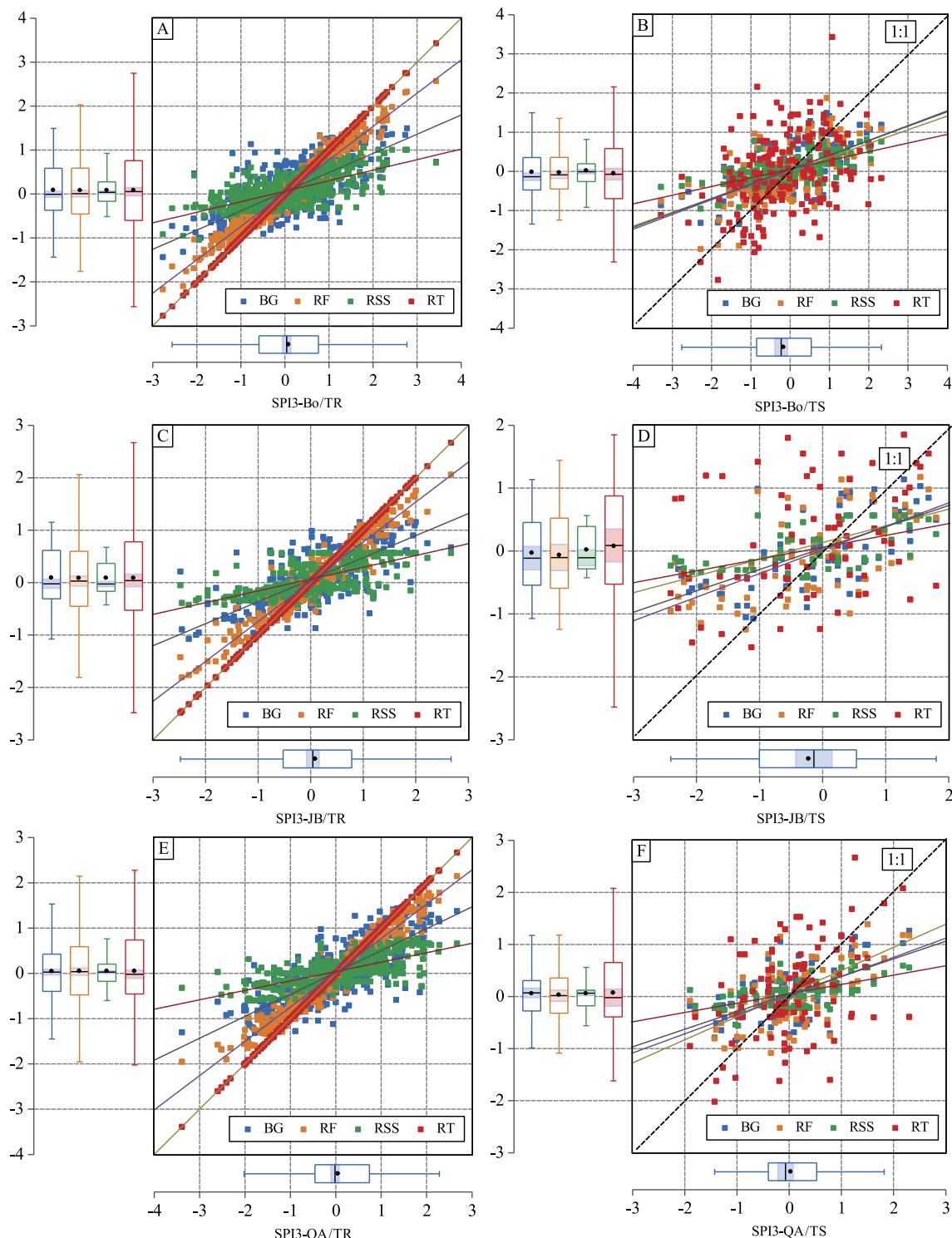
Among all the stations studied, the SN station showed a significant negative trend for all the drought scales, whereby the drought trend increased by 0.012 to 0.018 per decade. Some other stations such as BO, JB, JAB, and QA depicted a significant negative trend for SPI-3, -9, -12, and 24 (Fig. 4). However, the highest drought trend increase was recorded in JB, which ranged between 0.013 per decade (SPI-3) and 0.032 per decade (SPI-24).

### 3.2. Efficiency of machine learning methods for predicting drought events

#### 3.2.1. ML algorithms' performance in the training and testing stage

Three stations with different elevations were chosen for data mining training and testing. The stations were chosen based on their elevation to assess the ML algorithms' ability to capture different drought cycles. The first station was near the coast, namely BO (50 m), with an average rainfall of 836 mm (1946–2005). The second station was QA (750 m), with an average rainfall of 1278.92 mm (1959–2004), which was located mid-altitude of the coastal mountains. The JB was the highest climate station (950 mm), with an average rainfall of 1561.48 mm (1960–1990). After calculating drought indicators in different time scales, the SPI-3 was chosen as a representative of agricultural drought, while the SPI-12 was chosen as a representative of hydrological drought.

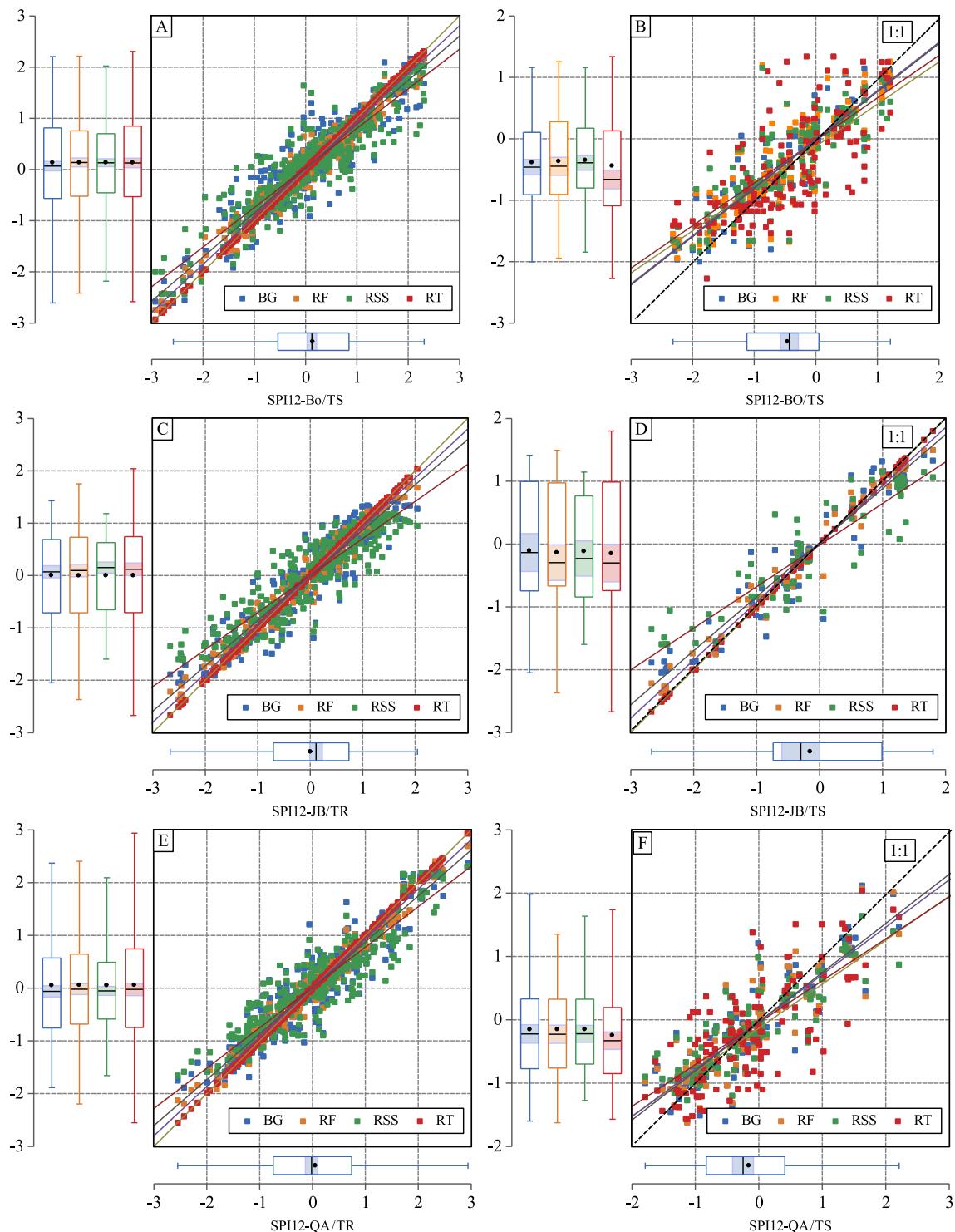
In the training stage, the RT algorithm showed a high aptitude for both agricultural and hydrological drought prediction (Fig. 5; Fig. 6).



**Fig. 5.** Observed values of SPI-3 (agricultural drought) and predicted one in three stations using four ML algorithms for the training (TR) and testing (TS) phases.

For the SPI-3, the statistical indicators showed that the RT algorithm had the lowest RMSE values, ranging between 0.3 for the BO station and 0.1 for both the JB and QA stations, and the highest  $r$  values, which range between 0.97 and 1 (Fig. 7A; Appendix 2a). Similar results were obtained for SPI-12, where RT and RF had the highest  $r$  values and the lowest RMSE (Fig. 7D; Appendix 3a). In this stage, the performance of the algorithms can be ranked as follows: RT > RF > BG > RSS for all stations and both drought indices, as can be seen in Fig. 8A, B, and C for SPI-3 and Fig. 9A, B, and C for SPI-12.

In the testing stage, BG, RSS, and RF showed dynamic preferences in drought prediction. Interestingly, the RT algorithm preference was the worst. The lowest RMSE value in the BO/SPI-3 station was recorded in BG accompanied with the lowest values for MAE (0.64), RAE (0.75), and RRSE (0.77); quite comparable results were also obtained for JB/SPI-3 and QA/SPI-3 (Fig. 7B; Appendix 2b). For other algorithms, both the RS and RF algorithms performed quite well (Fig. 7B; Appendix 2b), while the preference of RT was the lowest (Fig. 7B; Appendix 2b). For SPI-12, both the BG and RF algorithms had the highest correlation  $r$



**Fig. 6.** Observed values of SPI-12 (hydrological drought) and predicted one in three stations using four ML algorithms for the training (TR) and testing (TS) phases.

(observed vs. predicted) (0.58–0.64) and the lowest RMSE (0.68–0.88) (**Fig. 7E; Appendix 2b**). In contrast, the lowest correlation  $r$  (observed vs. predicted) (0.3–0.41) and highest RMSE (0.94–1.10) was calculated for the RT algorithm (**Fig. 7E; Appendix 2b**). The performance of the algorithms can be ranked as follows: BG > RF > RSS > RT for all stations and both drought indices, as can be seen in **Fig. 8D, E, and F** for SPI-3 and **Fig. 9D, E, and F** for SPI-12.

As both the BG and RF algorithms had the highest correlation and lowest RMSE, the Cr.V was implemented at 5 folds for the best both models (i.e., BG and RF). The results Cr.V were presented in

**Appendix 3.** Based on the output of the testing stage, the BG algorithm was chosen for validation.

### 3.2.2. ML algorithms' performance in the validation stage

As BG proved to have the highest performance, this algorithm was adopted and validated in the other three climate stations (LAT, SB, and TR), as depicted in **Fig. 10**. The validation climate stations were chosen based on geographical locations, where LAT (7 m) was located in the northern part of the study area (near the Turkish border), the TR (5 m) was located in the southern part of the study area (near the Lebanese

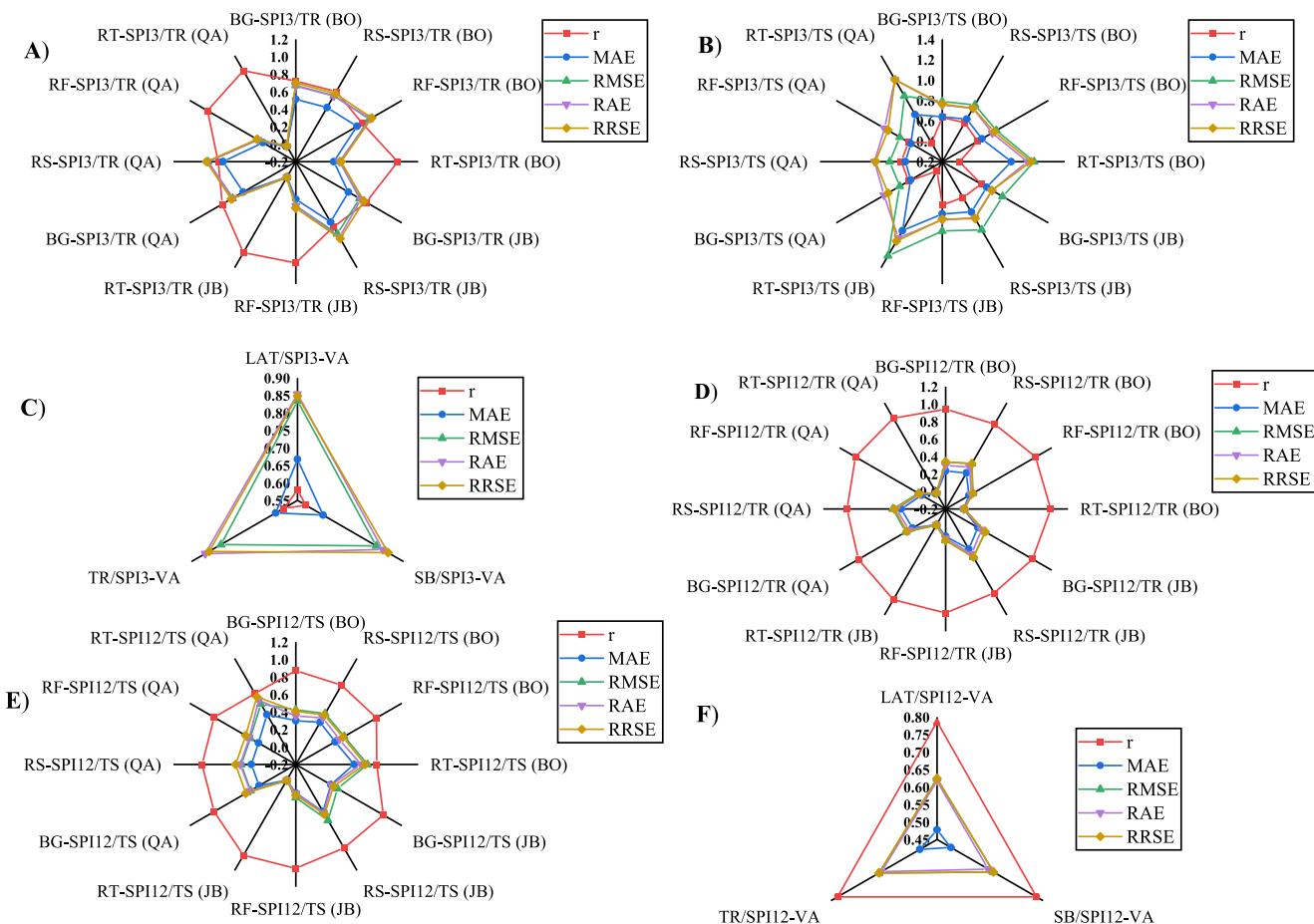


Fig. 7. Results of different statistical indices for evaluating the ML algorithms (training (TR), testing (TS), validation (VA)).

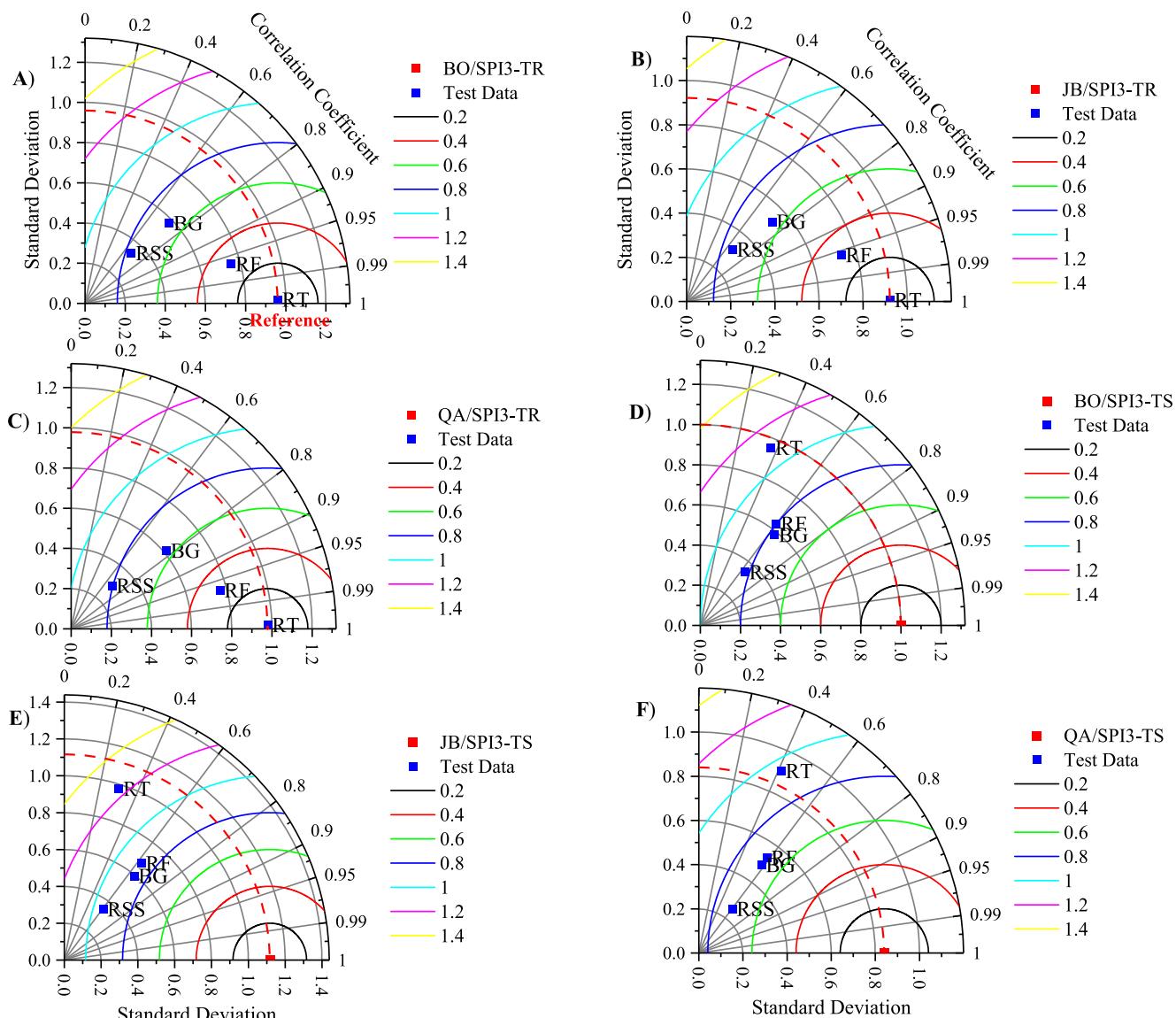
border), and SB (550 m) was chosen in the mountain region. In terms of the SPI-3 correlation,  $r$  (observed vs. predicted) ranged between 0.58 (LAT) and 0.6 (TR), while the RMSE did not exceed 0.83 (Fig. 7C; Appendix 2c). The performance of BG was powerful in drought production over a 12-month time scale (i.e., SPI-12) (Fig. 10), where the correlation  $r$  (observed vs. predicted) exceeded 0.75 for the three stations with a lower value of RMSE (Fig. 7F; Appendix 3c). Fig. 11 shows the relation between observed and predicted values for both SPI-3 and SPI-12 based on the Taylor diagram, where the highest compatibility was recorded in LAT-SPI-12 (Fig. 11D), while the lowest value in LAT-SPI-3 (Fig. 11A). Despite some variants in drought prediction across the station, BG seems to be a powerful algorithm in drought prediction in the eastern Mediterranean.

#### 4. Discussion:

##### 4.1. Drought trends across the eastern Mediterranean

The Mediterranean basin is one of the world's climate 'hotspots', where severe climate change consequences are projected (Giorgi, 2006). The Mediterranean is classified as a temperate mid-latitude climate with a hot summer season. Rainfall has a specific annual pattern, with little rainfall in summer and a regional gradient toward the south (Lionello, 2012). The transition phase between the warm and cold mild-latitudes causes seasonal and inter-annual variability in the Mediterranean climate, resulting in significant circulation shifts between winter and summer. The expected decrease in mean precipitation is one of the primary responses of the Mediterranean's climate to the rapid increase in greenhouse gas-driven warming (García-Ruiz et al., 2011; Cramer et al., 2018).

Over the last century, a drying trend in the Mediterranean region was recorded (Mariotti et al., 2015; Mathbou et al., 2021). In this study, both agricultural and hydrological drought events were captured across the CoR-SY (eastern Mediterranean) (Fig. 4). The output of this research coincides with previous research conducted in Syria. For instance, Mathbou et al. (2018) distinguished the CoR-SY as one of the regions that had a temporal evolution of droughts between 1960–2016. Mohammed et al. (2020) reported an increase in drought at SPI-12 (i.e., frequency and intensity) across Syria between 1990 and 2010. Due to poor records and high inter-annual precipitation variability, drought assessment in the Mediterranean basin is challenging. Drought events in this region are predicted to be frequent and severe in the future, according to both metrological (Dubrovský et al., 2014; Hertig and Tramblay, 2017) and hydrological (Forzieri et al., 2014) forecasts. Interestingly, Homsi et al. (2020) predicted future drier climate across Syria based on different RCPs pathways, which will affect the agricultural sector. The literature suggests various explanations for the Mediterranean dryness trend, including changes in the mean flow time on a large scale (Simpson et al., 2018; Seager et al., 2014), the expansion of the Hadley cell (Giorgi et al., 2019; Scheff and Frierson, 2012), and high-pressure anomalies (Seager et al., 2019; Giorgi and Lionello, 2008) or winter cyclone activity and weather regimes (Santos et al., 2016; Rojas et al., 2013; Giorgi and Lionello, 2008). Therefore, droughts in the eastern Mediterranean have been reported to be the most severe and long-lasting in decades (Kelley et al., 2015; Mathbou et al., 2021). Furthermore, several studies have revealed that the increased drought intensity in various Mediterranean countries (Jacob et al., 2018; Caloiero et al., 2018; Sousa et al., 2011; Raymond et al., 2019) is mainly due to the observed increase in atmospheric evaporative demand (Vicente-Serrano et al., 2014). On the other hand, drought events were projected



**Fig. 8.** Taylor diagram to compare observed and predicted (BG, RSS, RT, RF) drought events in a training (TR) and testing (TS) stages for three stations in terms of SPI-3.

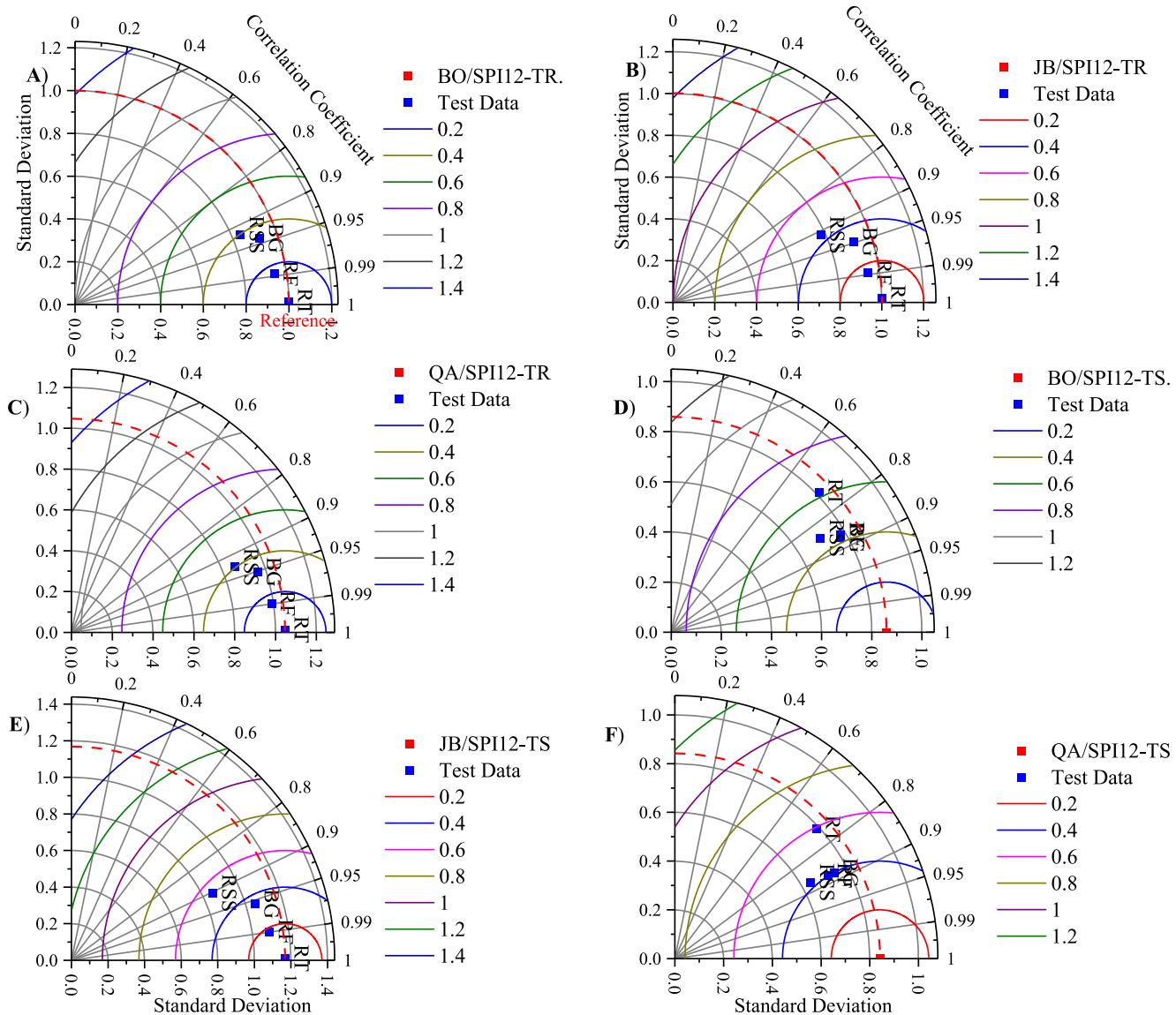
to become more frequent and severe across most of the world because of climate change (Cook et al., 2014).

#### 4.2. Efficiency of machine learning methods for predicting drought events

To build better adaptation methods, it is important to identify the possible implementation of machine learning algorithms and data mining techniques for future drought assessment. Scholars have recently demonstrated that some climatic phenomena such as drought events and their associated risks might be predicted using machine learning methods, which have high accuracy (Felsche and Ludwig, 2021; Park et al., 2016; Rahmati et al., 2020). Machine learning methods are now extensively being applied in a variety of scientific disciplines, such as flood prediction and assessment (Mandal and Pal, 2020), determining dust pollution (Ebrahimi-Khusfi et al., 2021), soil and landscape modeling (Zeraatpisheh et al., 2020), and landslide susceptibility valuation (Saha and Saha, 2020).

Previous studies have shown that machine learning models outperform traditional statistical methods. Additionally, machine learning algorithms can manage large datasets and result in more accurate findings (Gayen et al., 2019; Zeraatpisheh et al., 2020).

In this study, four machine learning algorithms (i.e., BG, RSS, RF, and RT) were implemented for the purpose of drought prediction. Despite the fact there is no unified approach for separating data for training and testing (Shamshirband et al. 2020). In this research, 70% of the dataset was used for model training, while 30% of the dataset was used for model testing. The results showed us that BG outperformed the other ML algorithms. One of the most fundamental advantages of the bagging technique is that it combines all the trees to produce a combined tree model rather than a single tree model output. It also removes the existing instability in the regression tree development. This was accomplished by eliminating the initial training datasets rather than by sampling new training datasets for each time step. In this context, the BG technique has successfully been applied to several environmental studies (Hong et al., 2020; Saha et al., 2021) and revealed excellent prediction accuracy. Nevertheless, this model is rarely used for drought prediction (Saha et al., 2021). It is good to mention here that the performance of the RF model was almost similar to the BG model, where the  $r$  (observed vs. predicted) and other statistical indicators were in the same range (Fig. 7; Appendices 2 and 3). Despite the different techniques in both algorithms, RF provided a set of equally probable drought indicator realizations, and the disparities between the realizations may be used to



**Fig. 9.** Taylor diagram to compare observed and predicted (BG, RSS, RT, RF) drought events in a training (TR) and testing (TS) stages for three stations in terms of SPI-12.

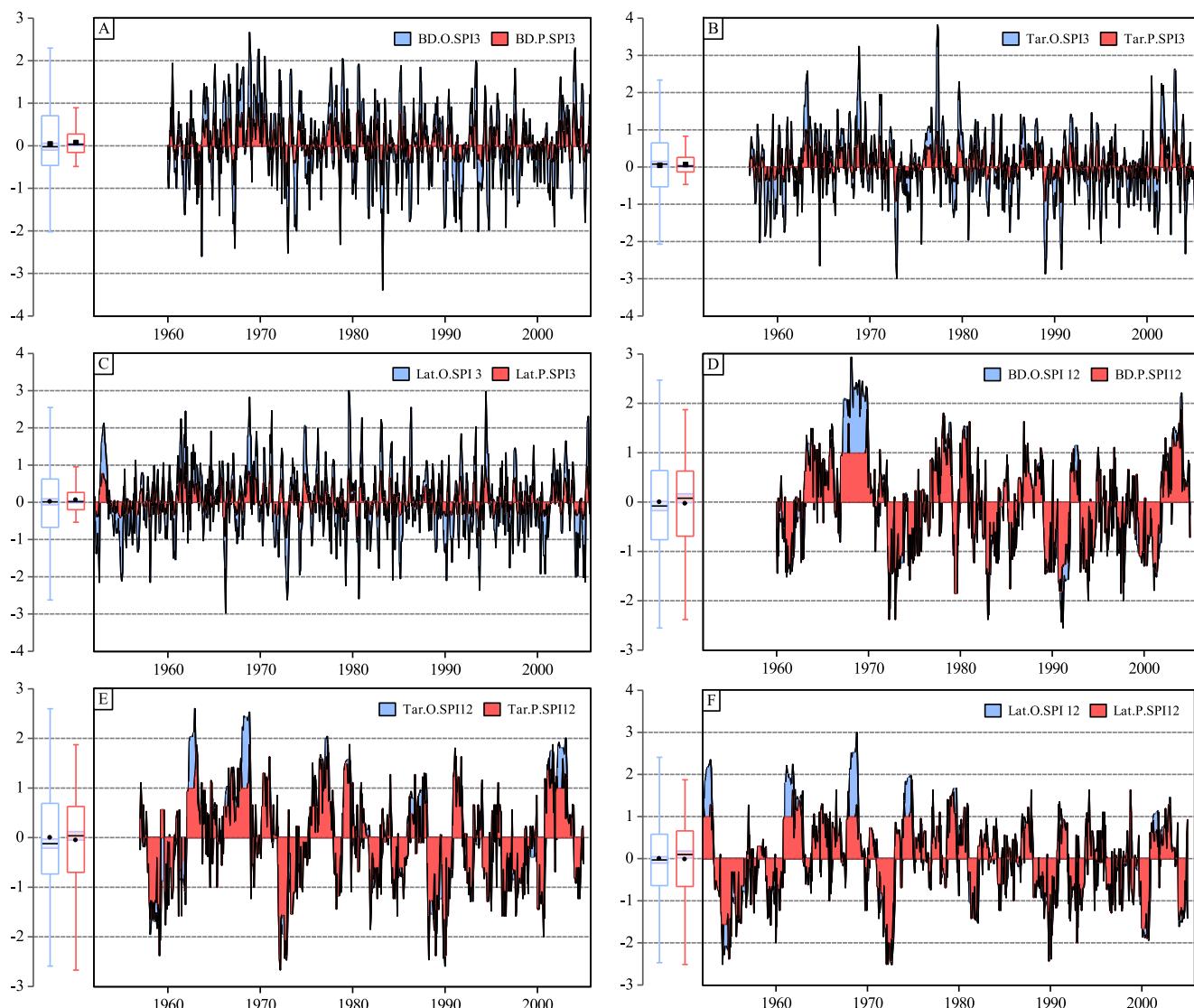
assess uncertainty (Chen et al., 2012; Park et al., 2016). Previous studies showed that the BG algorithm had a good drought prediction capacity (Saha et al., 2021a, b). On the other hand, RF is also reported to be one of the pioneer algorithms in drought prediction and assessment (Mokhtar et al., 2021; Danandeh Mehr et al., 2020).

#### 4.3. Limitation, uncertainty, and outlook

To capture drought across the Cor-SY, the SPI was implemented. The SPI, like all other indices, has its shortcomings. For instance, the SPI employs only rainfall data and did not consider the rest of the climatic factors (i.e., temperature, evapotranspiration, radiation). Also, the SPI not being able to forecast when a drought will begin and finish (Alsafadi et al., 2020). Some other features that could affect drought evolution were also did not consider by SPI, such as geographical and topographical differences. This research wholly relied on the results of the SPI index, which may affect prolonged drought over the study area. However, the SPI is a good index for drought monitoring that was

recommended by the World Meteorological Organization (WMO) and implemented worldwide.

On the other hand, selecting the appropriate input to predict hydrological events is a real challenge (Mohamadi et al. 2020). The principal component analysis is suggested to discover the best input combinations for modeling hydrological variables (Mohamadi et al. 2020). However, the output of the SPI index was used as input for the training and testing of the ML algorithms. Some uncertainty was revealed due to the nature of the input data, where the ML algorithms were not used for drought calculation but rather to simulate drought in the study area. According to Rahmati et al. (2020) and Swain et al. (2011), monitoring drought based on weather stations and interpolating the findings has a significant level of uncertainty in interpolated regions. Adopting a set of inputs (such as rainfall, temperature, and relative humidity) for drought simulation may improve the performance of various algorithms in capturing drought events. At a later stage, the SPEI could be compared with the SPI, considering a set of climatic factors to ensure the effectiveness of BG compared with the rest of the algorithms.



**Fig. 10.** Observed drought (blue line) and predicted one (red line) using the BG algorithm in three stations for two different time scales (SPI-3, SPI-12) (validation stage). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 5. Conclusion

In this study, drought prediction using four ML algorithms has been investigated in the Eastern Mediterranean region. Giving the fact that this region is suffering significantly from the changes in the climate, this effect was evidenced in the results of the initially drought trend analysis in this study. By using the SPI index, an increasing of drought trend over the region was detected, where the majority of the stations exhibited a significant negative ( $p < 0.05$ ) trend for both hydrological and agricultural drought.

ML algorithms (BG, RSS, RF, and RT) were then developed to test their performances in drought prediction. The performance of the ML algorithms varied depending on the implementation stage. For instance, the RT algorithm outperformed the other algorithms in the training stage, while both BG and RF outperformed the other algorithms in the testing stage. However, BG showed satisfactory performance for drought prediction when validated in three new climate stations across the CoR-

SY. According to the results of this study, using ML algorithms in drought research may improve prediction accuracy. In future work, the drought cycle prediction should be enhanced by using additional climate components, for example, evapotranspiration and soil moisture. This research provides the basis for applying ML methods to drought forecasting in the eastern Mediterranean in the future.

## CRediT authorship contribution statement

**Safwan Mohammed:** Conceptualization, Writing – original draft, Methodology. **Ahmed Elbeltagi:** Methodology. **Bashar Bashir:** Writing – review & editing. **Karam Alsafadi:** Methodology, Writing – review & editing. **Firas Alsilibe:** Writing – original draft. **Abdullah Alsalmam:** . **Mojtaba Zeraatpisheh:** Writing – review & editing. **Adrienn Széles:** Writing – review & editing. **Endre Harsányi:** Writing – review & editing.

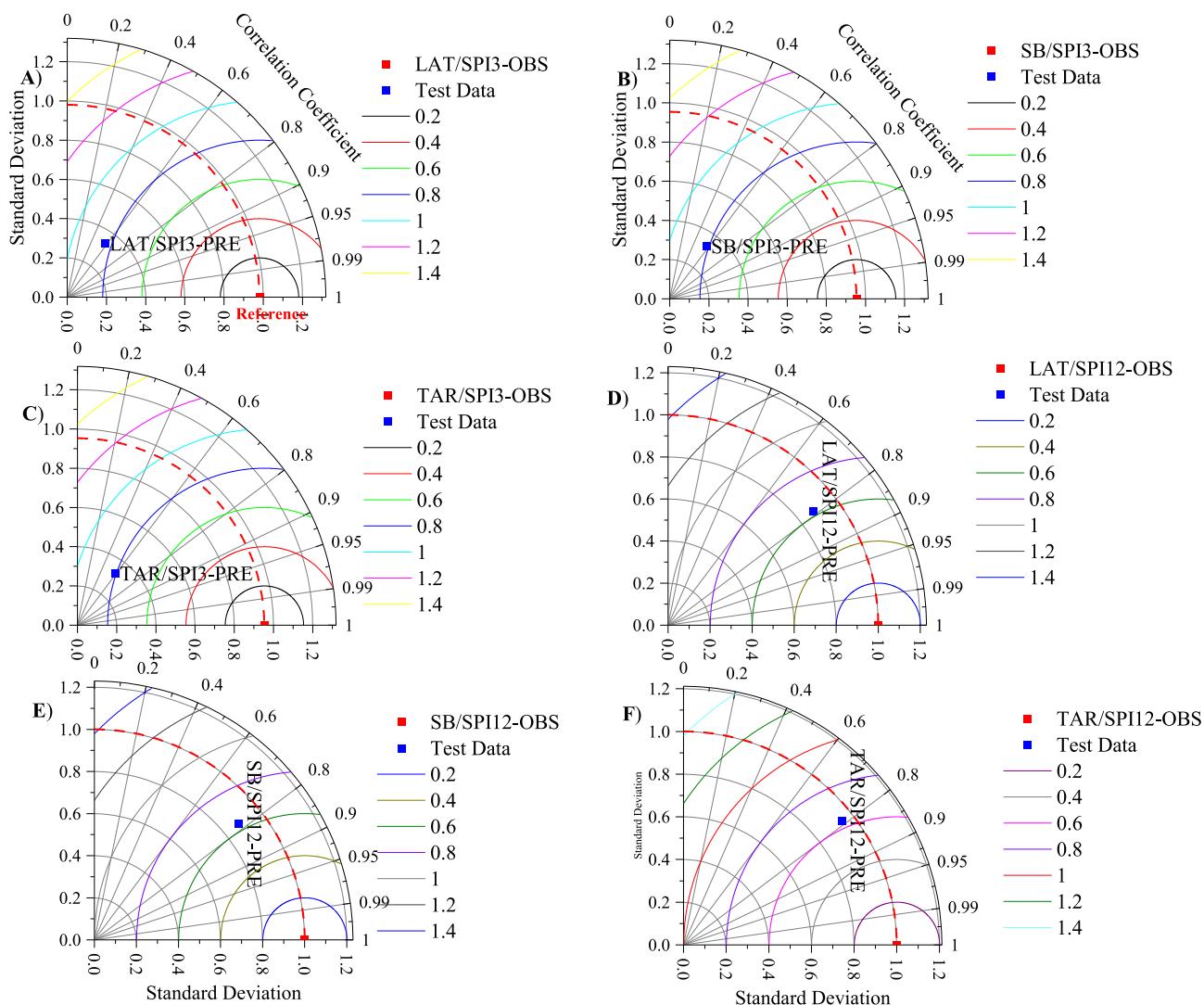


Fig. 11. Taylor diagram to compare observed and predicted (BG) drought events in a validation (VA) stage for three stations in terms of SPI-3, and SPI-12.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgement

Authors would like to thank **Debrecen university** for supporting open access. Also, authors would like to thank **Ministry of agriculture,**

**Damascus, Syria for providing the data.**

#### Funding

This research was supported by RESEARCHERS SUPPORTING PROJECT, grant number RSP-2021/296, King Saud University, Riyadh, Saudi Arabia. Also, Project no. TKP2021-NKTA-32 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme

#### Appendix 1. MK test and Sen slope over the study area

Station	Parameter	SPI-1	SPI-3	SPI-9	SPI-12	SPI-24
Arida	p	0.2810	0.2236	0.0741	0.0065	0.0014
	sen	-0.0002	-0.0003	-0.0006	-0.0008	-0.001
Banias	p	0.5394	0.3869	0.4914	0.7470	0.1349
	sen	0	-0.0003	-0.0004	-0.0002	-0.0006
Bouka	p	0.1275	0.0086	< 0.0001	< 0.0001	< 0.0001
	sen	-0.0002	-0.0005	-0.0008	-0.0009	-0.0011
Hmiemim	p	0.9127	0.7352	0.1802	0.1565	0.0365
	sen	0	0.0001	0.0006	0.0006	0.0008
Joubet Bourghal	p	0.1177	0.0066	< 0.0001	< 0.0001	< 0.0001

(continued on next page)

(continued)

Station	Parameter	SPI-1	SPI-3	SPI-9	SPI-12	SPI-24
Jablh	sen	-0.0006	-0.0013	-0.002	-0.0023	-0.0032
	p	0.0894	0.0335	0.0153	0.0081	0.0027
Lattakia	sen	-0.0011	-0.0019	-0.0022	-0.0025	-0.0025
	p	0.4247	0.3517	0.1298	0.0362	0.0005
Mashta Helo	sen	-0.0001	-0.0002	-0.0003	-0.0004	-0.0007
	p	0.8363	0.4073	0.1965	0.0813	< 0.0001
Qadmous	sen	0	-0.0002	-0.0003	-0.0005	-0.0012
	p	0.2861	0.0293	0.0024	0.0010	< 0.0001
Qirdaha	sen	-0.0002	-0.0005	-0.0009	-0.0009	-0.0014
	p	0.3095	0.0679	0.0015	0.0001	< 0.0001
Safita	sen	-0.0002	-0.0004	-0.0008	-0.001	-0.0015
	p	0.7830	0.6890	0.9706	0.7906	0.0836
Shaikh Bader	sen	0	0	0	0	-0.0005
	p	0.2861	0.0293	0.0024	0.0010	< 0.0001
Sin	sen	-0.0002	-0.0005	-0.0009	-0.0009	-0.0014
	p	0.0045	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Talkalakh	sen	-0.0006	-0.0012	-0.0014	-0.0015	-0.0018
	p	0.8891	0.8187	0.7784	0.8795	0.0634
Tartous	sen	0	0	-0.0001	0	-0.0008
	p	0.8988	0.7216	0.6939	0.7717	0.9009
	sen	0	0.0002	0.0002	0	0

**Appendix 2. Results of different statistical indices for evaluating the ML algorithms in terms of SPI-3****A) Training stage**

STA	BO/SPI3-TR				JB/SPI3-TR				QA/SPI3-TR			
	STAT	BG	RSS	RF	RT	BG	RSS	RF	RT	BG	RSS	RF
r	0.72	0.72	0.68	0.97	0.73	0.66	0.96	1.00	0.77	0.69	0.97	1.00
MAE	0.52	0.52	0.61	0.23	0.49	0.60	0.23	0.00	0.49	0.63	0.24	0.00
RMSE	0.67	0.67	0.77	0.30	0.64	0.75	0.31	0.01	0.64	0.80	0.30	0.01
RAE	0.67	0.67	0.79	0.30	0.66	0.80	0.31	0.00	0.63	0.80	0.30	0.00
RRSE	0.70	0.70	0.80	0.32	0.70	0.82	0.33	0.01	0.65	0.82	0.31	0.01

**B) Testing stage**

STAT	BO/SPI3-TS				JB/SPI3-TS				QA/SPI3-TS			
	BG	RSS	RF	RT	BG	RSS	RF	RT	BG	RSS	RF	RT
r	0.63	0.64	0.60	0.37	0.64	0.61	0.62	0.30	0.58	0.60	0.58	0.41
MAE	0.64	0.68	0.65	0.88	0.70	0.77	0.71	0.98	0.56	0.56	0.56	0.73
RMSE	0.79	0.84	0.81	1.10	0.88	0.97	0.88	1.27	0.68	0.71	0.68	0.94
RAE	0.75	0.80	0.76	1.03	0.76	0.84	0.77	1.06	0.86	0.86	0.85	1.13
RRSE	0.77	0.80	0.79	1.07	0.76	0.84	0.77	1.10	0.81	0.85	0.81	1.13

**C) Validation stage**

	LAT/SPI3-VA	SB/SPI3-VA	TR/SPI3-VA
r	0.58	0.58	0.60
MAE	0.67	0.63	0.62
RMSE	0.83	0.81	0.80
RAE	0.85	0.83	0.85
RRSE	0.85	0.85	0.84

**Appendix 3. Results of different statistical indices for evaluating the ML algorithms in terms of SPI-12****A) Training stage**

STAT	BO/SPI12-TR				JB/SPI12-TR				QA/SPI12-TR			
	BG	RSS	RF	RT	BG	RSS	RF	RT	BG	RSS	RF	RT
r	0.94	0.92	0.99	1.00	0.95	0.91	0.99	1.00	0.95	0.93	0.99	1.00
MAE	0.23	0.28	0.11	0.01	0.23	0.33	0.11	0.01	0.24	0.31	0.12	0.01
RMSE	0.34	0.40	0.16	0.01	0.32	0.44	0.16	0.01	0.33	0.40	0.16	0.01
RAE	0.30	0.35	0.14	0.01	0.27	0.40	0.13	0.01	0.28	0.36	0.14	0.01
RRSE	0.34	0.40	0.16	0.01	0.32	0.44	0.16	0.01	0.31	0.39	0.15	0.01

## B) Testing stage

STAT	BO/SPI12-TS				JB/SPI12-TS				QA/SPI12-TS			
	BG	RSS	RF	RT	BG	RSS	RF	RT	BG	RSS	RF	RT
r	0.87	0.85	0.87	0.73	0.96	0.90	0.99	1.00	0.88	0.87	0.88	0.74
MAE	0.30	0.36	0.32	0.47	0.26	0.42	0.13	0.01	0.29	0.31	0.29	0.46
RMSE	0.43	0.47	0.44	0.62	0.35	0.54	0.18	0.01	0.40	0.42	0.40	0.60
RAE	0.35	0.42	0.37	0.55	0.27	0.44	0.13	0.01	0.40	0.44	0.41	0.65
RRSE	0.41	0.45	0.42	0.59	0.30	0.46	0.15	0.01	0.46	0.49	0.46	0.69

## C) Validation stage

	LAT/SPI12-VA	SB/SPI12-VA	TR/SPI12-VA
r	0.79	0.78	0.78
MAE	0.48	0.50	0.51
RMSE	0.62	0.64	0.64
RAE	0.62	0.62	0.63
RRSE	0.62	0.64	0.64

## Appendix 3. Results of different statistical indices for cross-validation (Cr.V) for BG and RF

A) SPI-3		BO/SPI3- Cr.V		JB/SPI3- Cr.V		QA/SPI3- Cr.V	
STAT		BG	RF	BG	RF	BG	RF
r	0.54	0.54	0.51	0.49	0.57	0.54	
MAE	0.62	0.62	0.62	0.62	0.62	0.64	
RMSE	0.80	0.79	0.79	0.81	0.80	0.82	
RAE	80.70	80.10	82.60	83.05	79.29	81.52	
RRSE	83.63	82.92	85.61	88.02	81.86	84.02	

## B)SPI-12

B)SPI-12		BO/SPI12- Cr.V		JB/SPI12- Cr.V		QA/SPI12- Cr.V	
STAT		BG	RF	BG	RF	BG	RF
r	0.90	0.90	0.90	0.90	0.90	0.91	0.91
MAE	0.28	0.28	0.29	0.29	0.30	0.30	0.30
RMSE	0.42	0.43	0.41	0.42	0.41	0.42	0.42
RAE	36.36	36.72	35.34	35.94	36.42	36.45	
RRSE	42.57	43.06	41.79	42.07	39.23	40.29	

## References

- Ahmed, K., Shahid, S., Harun, S.B., Wang, X.-J., 2016. Characterization of seasonal droughts in Balochistan Province, Pakistan. Stoch. Env. Res. Risk Assess. 30 (2), 747–762.
- Ali, Z., Hussain, I., Faisal, M., Nazir, H.M., Hussain, T., Shad, M.Y., Mohamad Shoukry, A., et al., 2017. Forecasting Drought Using Multilayer Perceptron Artificial Neural Network Model. Advances in Meteorology 2017, 1–9.
- Alsaifadi, K., Mohammed, S.A., Ayugi, B., Sharaf, M., Harsányi, E., 2020. Spatial-temporal evolution of drought characteristics over Hungary between 1961 and 2010. Pure Appl. Geophys. 177 (8), 3961–3978.
- Amit, Y., Geman, D., 1997. Shape Quantization and Recognition with Randomized Trees. Neural Comput. 9 (7), 1545–1588.
- Angelidis, P., Maris, F., Kotsovinos, N., Hrissanthou, V., 2012. Computation of drought index SPI with alternative distribution functions. Water Resour. Manage. 26 (9), 2453–2473.

- Barua, S., Ng, A.W.M., Perera, B.J.C., 2012. Artificial neural network-based drought forecasting using a nonlinear aggregated drought index. *J. Hydrol. Eng.* 17 (12), 1408–1413.
- Belayneh, A., Adamowski, J., Khalil, B., Quilty, J., 2016. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmos. Res.* 172–173, 37–47.
- Bourdin, D.R., Fleming, S.W., Stull, R.B., 2012. Streamflow modelling: a primer on applications, approaches and challenges. *Atmos. Ocean* 50 (4), 507–536.
- Bouras, E.H., Jarlan, L., Er-Raki, S., Balaghi, R., Amazirh, A., Richard, B., Khabba, S., 2021. Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco. *Remote Sensing* 13 (16), 3101.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24 (2), 123–140.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Wadsworth Int. Group 37, 237–251.
- Busch, J.R., Ferrari, P.A., Flesia, A.G., Fraiman, R., Grynberg, S.P., Leonardi, F., 2009. Testing statistical hypothesis on random trees and applications to the protein classification problem. *Journal of Applied Statistics* 3, 542–563.
- Caloiero, T., Veltri, S., Caloiero, P., Frustaci, F., 2018. Drought analysis in Europe and in the Mediterranean basin using the standardized precipitation index. *Water* 10 (8), 1043.
- Chaves, M.M., Maroco, J.P., Pereira, J.S., 2003. Understanding plant responses to drought—from genes to the whole plant. *Funct. Plant Biol.* 30 (3), 239–264.
- Chen, J., Li, M., Wang, W., 2012. Statistical uncertainty estimation using random forests and its application to drought forecast. *Math. Probl. Eng.* 2012, 1–13. <https://doi.org/10.1155/2012/915053>.
- Cook, B.I., Smeddon, J.E., Seager, R., Coats, S., 2014. Global warming and 21st century drying. *Clim. Dyn.* 43 (9–10), 2607–2627.
- Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.-P., Iglesias, A., Lange, M.A., Lionello, P., Llasat, M.C., Paz, S., Peñuelas, J., Snoussi, M., Toreti, A., Tsimplis, M.N., et al., 2018. Climate change and interconnected risks to sustainable development in the Mediterranean. *Nat. Clim. Change* 8 (11), 972–980.
- Deo, R.C., Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmos. Res.* 153, 512–525.
- Deo, R.C., Kisi, O., Singh, V.P., 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and MTree model. *Atmos. Res.* 184, 149–175.
- Deo, R. C., Salcedo-Sanz, S., Carro-Calvo, L., & Saavedra-Moreno, B. (2018). Drought prediction with standardized precipitation and evapotranspiration index and support vector regression models. In *Integrating disaster science and management* (pp. 151–174). Elsevier.
- Dubrovský, M., Hayes, M., Duce, P., Trnka, M., Svoboda, M., Zara, P., 2014. Multi-GCM projections of future drought and climate variability indicators for the Mediterranean region. *Reg. Environ. Change* 14 (5), 1907–1919.
- Durrani, I. H., Adnan, S., & Afتاب, S. M. (2018, September). Historical and future climatological drought projections over Quetta Valley, Balochistan, Pakistan. In *IOP Conference Series: Materials Science and Engineering* (Vol. 414, No. 1, p. 012043). IOP Publishing.
- Ebrahimi-Khusfi, Z., Taghizadeh-Mehrjardi, R., Roustaei, F., Ebrahimi-Khusfi, M., Mosavi, A.H., Heung, B., Soleimani-Sardo, M., Scholten, T., 2021. Determining the contribution of environmental factors in controlling dust pollution during cold and warm months of western Iran using different data mining algorithms and game theory. *Ecol. Ind.* 132, 108287. <https://doi.org/10.1016/j.ecolind.2021.108287>.
- Fahimi, F., Yaseen, Z.M., El-shafie, A., 2017. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. *Theor. Appl. Climatol.* 128 (3–4), 875–903.
- Felsche, E., Ludwig, R., 2021. Applying machine learning for drought prediction using data from a large ensemble of climate simulations. *Nat. Hazards Earth Syst. Sci. Discuss.* 1–20 <https://doi.org/10.5194/nhess-2021-110>.
- Feng, P., 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. *Agric. Syst.* 173, 303–316.
- Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., Bianchi, A., 2014. Ensemble projections of future streamflow droughts in Europe. *Hydrol. Earth Syst. Sci.* 18 (1), 85–108.
- Ganguli, P., Reddy, M.J., 2014. Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrol. Process.* 28 (19), 4989–5009.
- García-Ruiz, J.M., López-Moreno, J.I., Vicente-Serrano, S.M., Lasanta-Martínez, T., Beguería, S., 2011. Mediterranean water resources in a global change scenario. *Earth Sci. Rev.* 105 (3–4), 121–139.
- Gayen, A., Pourghasemi, H.R., Saha, S., Keesstra, S., Bai, S., 2019. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci. Total Environ.* 668, 124–138.
- Ginoux, P., Prospero, J.M., Torres, O., Chin, M., 2004. Long-term simulation of global dust distribution with the GOCART model: correlation with North Atlantic Oscillation. *Environ. Monit. Assess.* Software 19 (2), 113–128. [https://doi.org/10.1016/S1364-8152\(03\)00114-2](https://doi.org/10.1016/S1364-8152(03)00114-2).
- Gioia, A., Bruno, M.F., Totaro, V., Iacobellis, V., 2020. Parametric Assessment of Trend Test Power in a Changing Environment. *Sustainability* 12 (9), 3889.
- Giorgi, F., 2006. Climate change hot-spots. *Geophys. Res. Lett.* 33 (8) <https://doi.org/10.1029/2006GL025734>.
- Giorgi, F., Lionello, P., 2008. Climate change projections for the Mediterranean region. *Global Planet. Change* 63 (2–3), 90–104.
- Giorgi, F., Raffaele, F., Coppola, E., 2019. The response of precipitation characteristics to global warming from climate projections. *Earth Syst. Dyn.* 10 (1), 73–89.
- Granata, F., 2019. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric. Water Manag.* 217, 303–315.
- Hao, Z., Singh, V.P., Xia, Y., 2018. Seasonal drought prediction: advances, challenges, and future prospects. *Rev. Geophys.* 56 (1), 108–141. <https://doi.org/10.1002/2016RG000549>.
- Harsányi, E., Bashir, B., Alsilibe, F., Alsafadi, K., Alsalman, A., Széles, A., et al., 2021. Impact of Agricultural Drought on Sunflower Production across Hungary. *Atmosphere* 12 (10), 1339. <https://doi.org/10.3390/atmos12101339>.
- Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M., 2019. Supervised learning with quantum-enhanced feature spaces. *Nature* 567 (7747), 209–212. <https://doi.org/10.1038/s41586-019-0980-2>.
- He, W., Perez, O.F.L., Fok, S.L., Marsden, B.J., 2003. A New Specimen Geometry for the Iosipescu Shear Test. *Mater. Sci. Forum* 440–441, 77–84. <https://doi.org/10.4028/www.scientific.net/msf.440-441.77>.
- Hertig, E., Tramblay, Y., 2017. Regional downscaling of Mediterranean droughts under past and future climatic conditions. *Global Planet. Change* 151, 36–48.
- Ho, T.K., 1998. The Random Subspace Method for Constructing Decision Forests. *IEEE Trans. pattern Anal. Mach. Intell.* 20, 832–844.
- Hong, H., Liu, J., Zhu, A.-X., 2020. Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. *Sci. Total Environ.* 718, 137231. <https://doi.org/10.1016/j.scitotenv.2020.137231>.
- Homsi, R., Shiru, M.S., Shahid, S., Ismail, T., Harun, S.B., Al-Ansari, N., Chau, K.-W., et al., 2020. Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. *Engineering Applications of Computational Fluid Mechanics* 14 (1), 90–106.
- Jacob, D., Kotova, L., Teichmann, C., Sobolowski, S.P., Vautard, R., Donnelly, C., et al., 2018. Climate impacts in Europe under + 1.5 °C global warming. *Earth's Future* 6 (2), 264–285.
- Jasim, A.I., Awchi, T.A., 2020. Regional meteorological drought assessment in Iraq. *Arabian J. Geosci.* 13 (7), 1–16.
- Karavitis, C.A., Alexandris, S., Tsitsmelis, D.E., Athanasopoulos, G., 2011. Application of the standardized precipitation index (SPI) in Greece. *Water* 3 (3), 787–805.
- Kelley, C.P., Mohtadi, S., Cane, M.A., Seager, R., Kushnir, Y., 2015. Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proc. Natl. Acad. Sci.* 112 (11), 3241–3246.
- Kendall, M.G., 1975. Rank Correlation Methods. Griffin, London.
- Khan, B., Naseem, R., Muhammad, F., Abbas, G., Kim, S., 2020. An empirical evaluation of machine learning techniques for chronic kidney disease prophecy. *IEEE Access* 8, 55012–55022.
- Khan, N., Shahid, S., Juneng, L., Ahmed, K., Ismail, T., Nawaz, N., 2019. Prediction of heat waves in Pakistan using quantile regression forests. *Atmos. Res.* 221, 1–11.
- Kreft, H., Jetz, W., 2007. Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci.* 104 (14), 5925–5930.
- Lionello, P. (Ed.). (2012). *The climate of the Mediterranean region: From the past to the future*. Elsevier.
- Lorenzo-Lacruz, J., García, C., Morán-Tejeda, E., 2017. Groundwater level responses to precipitation variability in Mediterranean insular aquifers. *J. Hydrol.* 552, 516–531.
- Lorenzo-Lacruz, J., Morán-Tejeda, E., Vicente-Serrano, S.M., López-Moreno, J.I., 2013. Streamflow droughts in the Iberian Peninsula between 1945 and 2005: spatial and temporal patterns. *Hydrol. Earth Syst. Sci.* 17 (1), 119–134.
- Madagari, S., Aghakouchak, A., Shukla, S., Wood, A.W., Cheng, L., Hsu, K.-L., Svoboda, M., 2016. A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States. *Water Resour. Res.* 52 (7), 5095–5110.
- Malik, A., Tikhamarine, Y., Sammen, S.S., Abba, S.I., Shahid, S., 2021a. Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. *Environ. Sci. Pollut. Res.* 28 (29), 39139–39158.
- Malik, A., Tikhamarine, Y., Souag-Gamane, D., Rai, P., Sammen, S.S., Kisi, O., 2021b. Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. *Meteorol. Atmos. Phys.* 133 (3), 891–909.
- Mandal, I., Pal, S., 2020. Modelling human health vulnerability using different machine learning algorithms in stone quarrying and crushing areas of Dwarka river Basin. *Eastern India. Adv. Space Res.* 66 (6), 1351–1371.
- Mann, H.B., 1945. Non-parametric tests against trend. *Econometrica: Journal of the Econometric Society* 13 (3), 245. <https://doi.org/10.2307/1907187>.
- Mariotti, A., Pan, Y., Zeng, N., Alessandri, A., 2015. Long-term climate change in the Mediterranean region in the midst of decadal variability. *Clim. Dyn.* 44 (5–6), 1437–1456.
- Mariotti, A., Schubert, S., Mo, K., Peters-Lidard, C., Wood, A., Pulwarty, R., Huang, J., et al., 2013. Advancing drought understanding, monitoring, and prediction. *Bull. Am. Meteorol. Soc.* 94 (12), ES186–ES188.
- Mathbout, S., Lopez-Bustins, J.A., Martin-Vide, J., Bech, J., Rodrigo, F.S., 2018. Spatial and temporal analysis of drought variability at several time scales in Syria during 1961–2012. *Atmos. Res.* 200, 153–168.
- Mathbout, S., Lopez-Bustins, J.A., Royé, D., Martin-Vide, J., 2021. Mediterranean-Scale Drought: Regional Datasets for Exceptional Meteorological Drought Events during 1975–2019. *Atmosphere* 12 (8), 941.
- McKee, T.B., Doesken, N.J., & Kleist, J. (1993) The relationship of drought frequency and duration to time scale. Preprints, Eighth Conf. on Applied Climatology, Anaheim, CA, Amer Meteor Soc 179–184.
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993, January). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (Vol. 17, No. 22, pp. 179–183).

- Danandeh Mehr, A., Tur, R., Çalışkan, C., Tas, E., 2020. A novel fuzzy random forest model for meteorological drought classification and prediction in ungauged catchments. *Pure Appl. Geophys.* 177 (12), 5993–6006.
- Mishra, A.K., Desai, V.R., 2006. Drought forecasting using feed-forward recursive neural network. *Ecol. Model.* 198 (1–2), 127–138.
- Mishra, A.K., Singh, V.P., 2011. Drought modeling—A review. *J. Hydrol.* 403 (1–2), 157–175.
- Mohammed, S., Alsafadi, K., Al-Awadhi, T., Sherief, Y., Harsanyie, E., El Kenawy, A.M., 2020. Space and time variability of meteorological drought in Syria. *Acta Geophys.* 68 (6), 1877–1898.
- Mohammed, S., Alsafadi, K., Mousavi, S. M. N., & Harsányi, E. (2021). Rainfall Change and Spatial-Temporal Aspects of Agricultural Drought in Syria. In *Water Resources in Arid Lands: Management and Sustainability* (pp. 215–221). Springer, Cham.
- Mohamadi, S., Sammen, S.S., Panahi, F., Ehteram, M., Kisi, O., Mosavi, A., et al., 2020. Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. *Nat. Hazards* 104 (1), 537–579.
- Mokhtar, A., He, H., Alsafadi, K., Mohammed, S., Ayantobo, O.O., Elbeltagi, A., , et al. Li, Y., 2021. Assessment of the effects of spatiotemporal characteristics of drought on crop yields in Southwest China. *Int. J. Climatol.*
- Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., et al., 2021. Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms. *IEEE Access* 9, 65503–65523.
- Mokhtarzad, M., Eskandari, F., Vanjani, N.J., Arabasadi, A., 2017. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environmental earth sciences* 76 (21), 1–10.
- Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 27 (15), 2103–2111.
- Mouatadid, S., Raj, N., Deo, R.C., Adamowski, J.F., 2018. Input selection and data-driven model performance optimization to predict the Standardized Precipitation and Evaporation Index in a drought-prone region. *Atmos. Res.* 212, 130–149.
- Murakami, H., Villarini, G., Vecchi, G.A., Zhang, W., Gudgel, R., 2016. Statistical-dynamical seasonal forecast of North Atlantic and US landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Mon. Weather Rev.* 144 (6), 2101–2123.
- Najock, D., Heyde, C.O., 1982. The number of terminal vertices in certain random trees with an application to stemma construction in philology. *Journal of Applied Probability* 19, 675–680.
- Nourani, V., Pradhan, B., Ghaffari, H., Sharifi, S.S., 2014. Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. *Nat. Hazards* 71 (1), 523–547.
- Park, S., Im, J., Jang, E., Rhee, J., 2016. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agric. For. Meteorol.* 216, 157–169. <https://doi.org/10.1016/j.agrmet.2015.10.011>.
- Parmar, A., Mistree, K., & Sompura, M. (2017, March). Machine learning techniques for rainfall prediction: A review. In *International Conference on Innovations in Information Embedded and Communication Systems* (Vol. 3).
- Pearson, K., 1896. Mathematical contributions to the theory of evolution: 111. Regression, heredity, and panmixia. *Philos. Trans. R. Soc. Lond. Ser. A* 187, 253–318.
- Peters, J., De Baets, B., Verhoest, N.E., Samson, R., Degroeve, S., De Becker, P., Huybrechts, W., 2007. Random forests as a tool for ecohydrological distribution modeling. *Ecol. Model.* 207, 304–318.
- Poornima, S., Pushpalatha, M., 2019. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft. Comput.* 23 (18), 8399–8412.
- Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J.V., et al., 2013. Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting. *Bull. Am. Meteorol. Soc.* 94 (6), 776–785.
- Rahmati, O., Falah, F., Dayal, K.S., Deo, R.C., Mohammadi, F., Biggs, T., Moghaddam, D. D., Naghibi, S.A., Bui, D.T., 2020. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Sci. Total Environ.* 699, 134230. <https://doi.org/10.1016/j.scitotenv.2019.134230>.
- Raymond, F., Ullmann, A., Camberlin, P., Drobinski, P., Smith, C.C., 2016. Extreme dry spell detection and climatology over the Mediterranean Basin during the wet season. *Geophys. Res. Lett.* 43 (13), 7196–7204.
- Raymond, F., Ullmann, A., Tramblay, Y., Drobinski, P., Camberlin, P., 2019. Evolution of Mediterranean extreme dry spells during the wet season under climate change. *Reg. Environ. Change* 19 (8), 2339–2351.
- Rhee, J., Im, J., 2017. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agric. For. Meteorol.* 237–238, 102–122.
- Rojas, M., Li, L.Z., Kanakidou, M., Hatzianastassiou, N., Seze, G., Le Treut, H., 2013. Winter weather regimes over the Mediterranean region: their role for the regional climate and projected changes in the twenty-first century. *Clim. Dyn.* 41 (3–4), 551–571.
- Sachindra, D.A., Kanae, S., 2019. Machine learning for downscaling: the use of parallel multiple populations in genetic programming. *Stoch. Env. Res. Risk Assess.* 33 (8–9), 1497–1533.
- Saha, A., Saha, S., 2020. Comparing the Efficiency of Weight of Evidence, Support Vector Machine and Their Ensemble Approaches in Landslide Susceptibility Modelling: A Study on Kurseong Region of Darjeeling Himalaya, India. *Remote Sens. Appl.: Soc. Environ.* 19, 100323. <https://doi.org/10.1016/j.rsase.2020.100323>.
- Saha, S., Gayen, A., Gogoi, P., Kundu, B., Paul, G.C., Pradhan, B., 2021a. Proposing novel ensemble approach of particle swarm optimized and machine learning algorithms for drought vulnerability mapping in Jharkhand, India. *Geocarto International* 1–32.
- Saha, S., Gogoi, P., Gayen, A., Paul, G.C., 2021b. Constructing the machine learning techniques based spatial drought vulnerability index in Karnataka state of India. *J. Clean. Prod.* 314, 128073. <https://doi.org/10.1016/j.jclepro.2021.128073>.
- Santos, J.A., Belo-Pereira, M., Fraga, H., Pinto, J.G., 2016. Understanding climate change projections for precipitation over western Europe with a weather typing approach. *Journal of Geophysical Research: Atmospheres* 121 (3), 1170–1189.
- Scheff, J., Frierson, D., 2012. Twenty-first-century multimodel subtropical precipitation declines are mostly midlatitude shifts. *J. Clim.* 25 (12), 4330–4347.
- Schilling, J., Hertig, E., Tramblay, Y., Scheffran, J., 2020. Climate change vulnerability, water resources and social implications in North Africa. *Reg. Environ. Change* 20 (1), 1–12.
- Seager, R., Liu, H., Henderson, N., Simpson, I., Kelley, C., Shaw, T., et al., 2014. Causes of increasing aridification of the Mediterranean region in response to rising greenhouse gases. *J. Clim.* 27 (12), 4655–4676.
- Seager, R., Osborn, T.J., Kushnir, Y., Simpson, I.R., Nakamura, J., Liu, H., 2019. Climate variability and change of Mediterranean-type climates. *J. Clim.* 32 (10), 2887–2915.
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* 63 (324), 1379–1389.
- Shahid, S., 2010. Rainfall variability and the trends of wet and dry periods in Bangladesh. *Int. J. Climatol.* 30 (15), 2299–2313.
- Sharma, A., Goyal, M.K., 2020. Assessment of drought trend and variability in India using wavelet transform. *Hydrol. Sci. J.* 65 (9), 1539–1554.
- Shamshirband, S., Hashemi, S., Salimi, H., Samadianfar, S., Asadi, E., Shadkani, S., et al., 2020. Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics* 14 (1), 339–350.
- Simpson, I.R., Hitchcock, P., Seager, R., Wu, Y., Callaghan, P., 2018. The downward influence of uncertainty in the Northern Hemisphere stratospheric polar vortex response to climate change. *J. Clim.* 31 (16), 6371–6391.
- Sousa, P.M., Trigo, R.M., Aizpurua, P., Nieto, R., Gimeno, L., Garcia-Herrera, R., 2011. Trends and extremes of drought indices throughout the 20th century in the Mediterranean. *Nat. Hazards Earth Syst. Sci.* 11 (1), 33–51.
- Strazzo, S., Collins, D.C., Schepen, A., Wang, Q.J., Becker, E., Jia, L., 2019. Application of a hybrid statistical-dynamical system to seasonal prediction of North American temperature and precipitation. *Mon. Weather Rev.* 147 (2), 607–625.
- Swain, S., Wardlow, B.D., Narumalani, S., Tadesse, T., Callahan, K., 2011. Assessment of vegetation response to drought in Nebraska using Terra-MODIS land surface temperature and normalized difference vegetation index. *GI Sci. Remote. Sens.* 48 (3), 432–455.
- Taghizadeh-Mehrjardi, R., Nabibollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region. *Iran. Geoderma* 266, 98–110.
- Teuling, A.J., Van Loon, A.F., Seneviratne, S.I., Lehner, I., Aubinet, M., Heinesch, B., et al., 2013. Evapotranspiration amplifies European summer drought. *Geophys. Res. Lett.* 40 (10), 2071–2075.
- Tian, Y.e., Xu, Y.-P., Wang, G., 2018. Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin. *Sci. Total Environ.* 622–623, 710–720.
- Tuel, A., Eltahir, E.A., 2020. Why is the Mediterranean a climate change hot spot? *J. Clim.* 33 (14), 5829–5843.
- Turco, M., Ceglar, A., Prodhomme, C., Soret, A., Toreti, A., Doblas-Reyes Francisco, J., 2017. Summer drought predictability over Europe: empirical versus dynamical forecasts. *Environ. Res. Lett.* 12 (8), 084006. <https://doi.org/10.1088/1748-9326/aa7859>.
- Turkes, M., Turp, M. T., An, N., Ozturk, T., & Kurnaz, M. L. (2020). Impacts of climate change on precipitation climatology and variability in Turkey. In *Water resources of Turkey* (pp. 467–491). Springer, Cham.
- Ulgen, K.O.R.A.Y., Hepbasli, A., 2002. Comparison of solar radiation correlations for Izmir, Turkey. *International Journal of Energy Research* 26 (5), 413–430.
- van Dijk, A.I.J.M., Beck, H.E., Crosbie, R.S., de Jeu, R.A.M., Liu, Y.Y., Podger, G.M., et al., 2013. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resour. Res.* 49 (2), 1040–1057. <https://doi.org/10.1002/wrcr.20123>.
- Vicente-Serrano, S.M., Lopez-Moreno, J.-I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J.M., et al., 2014. Evidence of increasing drought severity caused by temperature rise in southern Europe. *Environ. Res. Lett.* 9 (4), 044001. <https://doi.org/10.1088/1748-9326/9/4/044001>.
- Vicente-Serrano, S.M., Quiring, S.M., Peña-Gallardo, M., Yuan, S., Domínguez-Castro, F., 2020. A review of environmental droughts: Increased risk under global warming? *Earth Sci. Rev.* 201, 102953. <https://doi.org/10.1016/j.earscirev.2019.102953>.
- Vishwakarma, D.K., Pandey, K., Kaur, A., Kushwaha, N.L., Kumar, R., Ali, R., et al., 2022. Methods to estimate evapotranspiration in humid and subtropical climate conditions. *Agric. Water Manag.* 261, 107378. <https://doi.org/10.1016/j.agwat.2021.107378>.
- Wang, W.-C., Chau, K.-W., Cheng, C.-T., Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.* 374 (3–4), 294–306.
- Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil* 340 (1–2), 7–24.

- Wilhite, D.A., Svoboda, M.D., Hayes, M.J., 2007. Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resour. Manage.* 21 (5), 763–774.
- Wilhite, D., & Pulwarty, R. S. (Eds.). (2017). *Drought and water crises: integrating science, management, and policy*. CRC Press.
- Wu, H., Soh, L.-K., Samal, A., Chen, X.-H., 2008. Trend analysis of streamflow drought events in Nebraska. *Water Resour. Manage.* 22 (2), 145–164.
- Xu, L., Chen, N., Zhang, X., Chen, Z., 2018. An evaluation of statistical, NMME and hybrid models for drought prediction in China. *J. Hydrol.* 566, 235–249.
- Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J. Hydrol.* 530, 829–844.
- Zeraatpisheh, M., Jafari, A., Bagheri Bodaghabadi, M., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., Kerry, R., Xu, M., 2020. Conventional and digital soil mapping in Iran: Past, present, and future. *Catena* 188, 104424. <https://doi.org/10.1016/j.catena.2019.104424>.
- Zhang, J., Ding, J., Wu, P., Tan, J., Huang, S., Teng, D., Cao, X., Wang, J., Chen, W., 2020. Assessing arid Inland Lake Watershed Area and Vegetation Response to Multiple Temporal Scales of Drought Across the Ebinur Lake Watershed. *Sci Rep* 10 (1). <https://doi.org/10.1038/s41598-020-57898-8>.
- Zhao, Z., Liu, H., 2007. Searching for interacting features. Proc. of International Joint Conference on Artificial Intelligence (IJCAI). Hyderabad, India.