



Drought analysis using innovative trend analysis and machine learning models for Eastern Black Sea Basin

Majid Niazkar¹ · Reza Piraei² · Gökçen Eryılmaz Türkkân³ · Tuğçe Hırca⁴ · Fabiola Gangi⁵ · Seied Hosein Afzali²

Received: 30 August 2023 / Accepted: 15 October 2023 / Published online: 31 October 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2023

Abstract

This study aims to assess the Eastern Black Sea Basin drought conditions. For this purpose, the trend changes in SPI values of 6, 9, 12, and 24 months using innovative trend analysis were examined. Additionally, four machine learning models, including Multiple Linear Regression, Artificial Neural Networks, K Nearest Neighbors, and XGBoost Regressor, are employed to forecast SPI with rainfall data between 1965 and 2020 from eight rainfall stations. The input data for each model was SPI values from lead times of 1 to 6, resulting into 768 unique scenarios. The ML models estimated SPI values better as the SPI duration increased, with the 24-month SPI showing the highest accuracy. The results of SPI forecast indicated that the optimal model and number of input variables varied for each SPI and station, indicating that further studies are required to improve SPI predictions.

1 Introduction

Climate change is a global challenge that poses significant threats to human societies and ecosystems. Drought, as an extreme event, is a consequence of climate change and/or anthropogenic origins. It is also a recurring natural hazard that inevitably put pressure on water resource management, hydrological water cycle, agriculture, and ecosystem health. Accurate and reliable predictions of future climate scenarios are crucial for effective mitigation and adaptation measures. However, climate modelling and analysis are highly complex and involve numerous factors, such as atmospheric dynamics, ocean currents, and land surface interactions. In the parlance of mitigating natural hazard impacts, it is essential not only to study the drought frequency but also to forecast drought indices.

The Standardized Precipitation Index (SPI) is a widely used meteorological index of drought. Basically, it is closely related to soil moisture and the groundwater reserves (Spennemann et al. 2015). However, predicting drought conditions exclusively using SPI can be challenging due to the complex interactions between meteorological variables and the environment. Therefore, powerful tools are required to capture such interactions and available pattern between various parameters involved.

Machine learning (ML) models have been used as versatile estimation tools to achieve possibly high accuracy of

✉ Majid Niazkar
majid.niazkar@unibz.it

Reza Piraei
piraeir@gmail.com

Gökçen Eryılmaz Türkkân
gokcen.turkkan@balikesir.edu.tr

Tuğçe Hırca
tugcehirca1@gmail.com

Fabiola Gangi
fabiola.gangi@unimi.it

Seied Hosein Afzali
afzali@shirazu.ac.ir

¹ Faculty of Engineering, Free University of Bozen-Bolzano, Piazza Domenicani 3, 39100 Bolzano, Italy

² Department of Civil Engineering, Shiraz University, Shiraz 7134851156, Iran

³ Department of Civil Engineering, Balıkesir University, 10145 Balıkesir, Türkiye

⁴ Department of Civil Engineering, Bayburt University, 69010 Bayburt, Türkiye

⁵ Department of Agricultural and Environmental Sciences, University of Milan, Via Celoria 2, Milan, Italy

predictions. These techniques have been applied to various aspects of climate research, including climate modelling, impact assessment, and adaptation planning (Dikshit et al. 2022). To be more precise, they have shown promising results in improving accuracy and efficiency of climate models, enabling faster analysis of large datasets, and supporting the development of predictive models for future climate scenarios.

Based on the literature, some attempts have been made regarding application of ML models for estimating SPI. For instance, Almedeij (2014) used a wide range of monthly total rainfall data from January 1967 to December 2009. They utilized SPI series for intermediate- and long-time scales of 3, 6, 12, and 24 months. They performed univariate and bivariate frequency analyses for the drought events. Furthermore, Belayneh et al. (2014) compared the effectiveness of five data-driven models for forecasting long-term (6- and 12-month lead time) drought conditions in the Awash River Basin of Ethiopia. Additionally, Hosseini-Moghari and Araghinejad (2015) predicted SPI for a 35-year period from 1972 to 2006 in the Gorganroud basin using six approaches of neural networks. Moreover, Belayneh et al. (2016) explored the ability of coupled ML models and ensemble techniques to predict drought conditions in the Awash River Basin of Ethiopia. Also, Maca and Pech (2016) forecasted drought indices, i.e. SPI and the Standardized Precipitation Evaporation Index (SPEI), based on two different models of Artificial Neural Networks (ANN) for the period of 1948–2002 for two catchments in the USA. In addition, Mondol et al. (2017) studied the meteorological drought by SPI using the Inverse Distance Weighted method. They utilized rainfall data of 30 meteorological stations in Bangladesh from 1981 to 2010. Their results indicated that drought has been fluctuating and consequently becomes a recurrent phenomenon during the study period. Also, El Ibrahimy and Baali (2018) predicted drought by applying Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Artificial Neural Network of Multi-Layered Perceptron (ANN-MLP), and Support Vector Regression (SVR) models. Abeyasingha and Rajapaksha (2020) assessed the status of drought in Sri Lanka to estimate SPI at 3, 6, and 12 months for rainfall data between 1970 and 2017 over 54 weather stations. The frequency of drought events was evaluated using SPI, while the trend of SPI was also detected using Mann–Kendall's (MK) test and Sen's slope estimator. Furthermore, Khan et al. (2020) combined Autoregressive Integrated Moving Average (ARIMA) with ANN to develop a hybrid model to predict future droughts for Malaysia's Langat River Basin. A 30-year period of rainfall data from 1986 to 2016 was used. Additionally, Adikari et al. (2021) compared three Artificial Intelligence (AI) techniques for flood and drought forecasting. The former was measured by the change in river discharge, while the latter was explored by SPI. Moreover,

Malik et al. (2021) investigated AI models, comprising Multi-Layer Perceptron Neural Network (MLPNN) and Co-Active Neuro-Fuzzy Inference System (CANFIS), and a Multiple Linear Regression (MLR) for multi-scalar SPI predictions in the Garhwal region of Uttarakhand State, India. Additionally, Taylan et al. (2021) employed ANFIS, SVM, and ANN to estimate drought in Çanakkale, Türkiye, using wavelet transform (W) for 3-, 6-, 9-, and 12-month SPI between 1975 and 2010. Furthermore, Prodhan et al. (2022) evaluated changes in future projected drought metrics and the future risk of yield reduction under drought intensity. Also, Ghazipour and Mahjouri (2022) developed a Bayesian Maximum Entropy-based Fusion (BMEF) model to improve the results of seasonal drought forecasting. Moreover, Elbeltagi et al. (2023b) estimated SPI using Random Forest (RF), Random Tree (RT), and Gaussian Process Regression (GPR) in a semi-arid region. A different combination of ML models and variables has been tested for forecasting the meteorological drought based on the SPI 6 and 12 months for the period of 2000–2019 at two meteorological stations, which frequently experience droughts. Additionally, Elbeltagi et al. (2023a) examined the feasibility and effectiveness of the Random Subspace (RSS) model and its hybridization with the M5 Pruning tree (M5P), RF, and RT to estimate SPI at 3, 6, and 12 months during 2000–2019. Also, Pande et al. (2023) adopted additive regression, RSS, M5P, and bagging tree models to predict SPI at the Upper Godavari Basin for 3, 6, and 12 months. Finally, Vodounon and Soude (2022) utilized RF and XGBoost to predict 3-, 6-, and 12-month SPIs for Alibori department in Benin Republic. They concluded that XGBoost performed better than RF. However, instead of considering SPI values of previous lead times as input data, they took into account other inflecting variables, such as soil moisture, total rainfall, relative humidity, wind direction, and air temperature. In contrast, the present study employed XGBR based on SPI values of previous lead times, and to the best of the authors' knowledge, this is the first implementation of XGBR for drought analysis solely based on SPI data.

According to the literature review, there is indeed a growing interest in employing ML algorithms for climate change impact assessments and analysis of extreme events like drought. In line with previous efforts on ML-based drought analysis and forecast, there is still a need to explore the applicability of novel and different ML models for the purpose of achieving more robust SPI estimations. Such improvements can help in better estimation of drought forecast and understanding occurrence of hydrological extreme events in light of taking counter measurements for mitigating impacts of climate change and extreme events.

In this study, drought condition of a region in Türkiye was assessed. For this purpose, trend changes in SPI values for different time scales were investigated by the innovative

trend analysis (ITA) method. Then, four ML models including MLR, K Nearest Neighbors (KNN), XGBoost, and ANN models were utilized to forecast SPI for eight rainfall stations using different SPI lag times as inputs. Because of many case scenarios compared in this study, it is postulated that it not only provides a suitable perspective of the drought condition of the study area but also presents an adequate comparative analysis among different ML models for SPI estimations. Because of substantial differences in the core concepts of the ML models employed in this study, the findings and comparative analysis can play a significant role in facilitating the future application of ML models for SPI predictions.

2 Study area

Eastern Black Sea Basin (EBSB) is on the northeast coast of Türkiye (Fig. 1). The basin is between 40°15' and 41°34' north latitudes and 36°43' and 41°35' east longitudes and encompassed by the Eastern Black Sea Mountains in the south and the Black Sea in the North. The surface water potential of the basin has an average of 14.9 km³ per year (Yüksek et al. 2013). With the effect of topographic factors, rainfall rises from the east of Trabzon and becomes larger in Rize, Arhavi, and Hopa (Karstarlı et al. 2011). Generally, the average annual rainfall within the EBSB is 1045 mm for Ordu, 1288 mm for Giresun, 830 mm for Trabzon, 2304 mm for Rize, 693 mm for Artvin, 718 mm for Samsun and 462 mm for Gümüşhane, respectively (GDM

2020). In total, the basin has an average rainfall height of 753 mm per year. With this amount of rainfall, the share of the basin in the total rainfall of Türkiye has been determined as 9.5%, and with this feature, EBSB has an important water potential (Odemis and Evrendilek 2007). This is why the monthly rainfall data of 8 stations located in EBSB between 1965 and 2020 were used to investigate drought analysis in the present study. Table 1 also shows the location and some statistical characteristics of rainfall in 8 rainfall-monitoring stations in EBSB.

3 Methodology

3.1 Standardized Precipitation Index

SPI is one of the most popular drought indices and widely recognized for characterizing meteorological droughts suitable for different timescales (1, 3, 6, 12, 24, and 48 months). Furthermore, SPI drought conditions are (1) normal for $0.99 < \text{SPI} < -0.99$, (2) moderately dry for $-1.00 < \text{SPI} < -1.49$, (3) severely dry for $-1.50 < \text{SPI} < -1.99$, (4) extremely dry for $\text{SPI} < -2.00$, (5) extremely wet for $\text{SPI} > 2.00$, (6) very wet for $1.50 < \text{SPI} < 1.99$, and (7) moderately wet for $1.00 < \text{SPI} < 1.49$. To be more precise, negative SPI values imply lower than average rainfall, whereas positive values denote more than average rainfall. Since rainfall data may be fitted by a gamma distribution, SPI is calculated using a probability density function of the gamma distribution:

Fig. 1 Location map of the Eastern Black Sea Basin and Türkiye



Table 1 Location features and rainfall characteristics of the stations in the study area

Station number	Station name	Latitude	Longitude	Rainfall period	Min. precipitation values (mm)	Max. precipitation values (mm)	Average of precipitation values (mm)
1	Rize/Pazar	41°10'39.7"N	40°53'57.5"E	1965–2020	0.0	703.2	170.77
2	Akçaabat	41°01'57.0"N	39°33'41.4"E	1965–2020	0.0	231.7	59.20
3	Giresun	40°55'21.7"N	38°23'16.1"E	1965–2020	0.2	521.6	103.63
4	Gümüşhane	40°27'35.3"N	39°27'55.1"E	1965–2020	0.0	141.9	37.46
5	Hopa	41°24'23.4"N	41°25'58.8"E	1965–2020	9.3	709.4	183.39
6	Ordu	40°59'01.7"N	37°53'08.9"E	1965–2020	0.3	251.0	86.12
7	Rize	41°02'24.0"N	40°30'04.7"E	1965–2020	6.8	516.6	183.49
8	Ünye	41°08'34.8"N	37°17'34.8"E	1981–2020	0.0	371.2	98.12

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \tau(\alpha)} x^{\alpha-1} e^{-x/\beta} \text{ for } x, \alpha, \beta > 0 \tag{1}$$

where β , α , x , and $\tau(\alpha)$ represent the scale, shape variables, rainfall amount, and gamma function, respectively. The finest values of α and β can be obtained by Eqs. (2) and (3), respectively (Guttman 1999):

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \tag{2}$$

$$\beta = \frac{\bar{x}}{\alpha} \tag{3}$$

where $A = \ln \bar{x} - \frac{\sum \ln(x)}{n}$, \bar{x} and n denote the average rainfall and number of observations, respectively.

3.1.1 SPI-based scenarios

In this study, SPI was calculated for four time scales, i.e. 6, 9, 12, and 24 months, using a MATLAB open source code (Taesam Lee 2023). The SPI prediction is a recursive process, which basically utilizes outcomes of preceding SPI as an input for future SPI predictions. Furthermore, input variables varied based on different lead times, which are ranging from 1 to 6. For instance, when considering a lead time of 3, input variables are SPIs at time $t-1$, $t-2$, and $t-3$, while the target variable is SPI at time t .

3.2 Innovative trend analysis method

ITA, which was first proposed by Şen (2012), graphically examines changes in hydrological and meteorological parameters. Unlike classical trend analysis methods, such as Mann–Kendall’s test and Spearman’s rho, it is not subject to constraints, like data length, independent structure of time series, and normality assumption. On the other hand, it is open to interpretation instead of the monotonous trend detection in classical methods (Şen 2012; Farrokhi

et al. 2020; Hirca and Eryılmaz Türkkan 2022). Because of these advantages, it is a popular trend determination method and consequently used to quantify changes in meteorological parameters (Dabanlı et al. 2016; Hirca et al. 2022) and drought indexes (Caloiero 2018; Yilmaz 2019).

ITA is basically based on marking the data in the Cartesian coordinate system (Fig. 2). For this purpose, the data length is mainly divided into 2 equal parts ordered from the smallest to the largest value. Afterwards, the first half is marked to the x -axis, whereas the second half is marked to the y -axis. The trend is interpreted according to the scattering of the data around the identical line on the graph. To be more precise, if it is in the upper triangle area above the identical line, it means that the data has an increasing trend. However, if it is in the lower triangle area below the identical line, it implies that it has a decreasing trend. Finally, in case it is above the line, it does not have any trends.

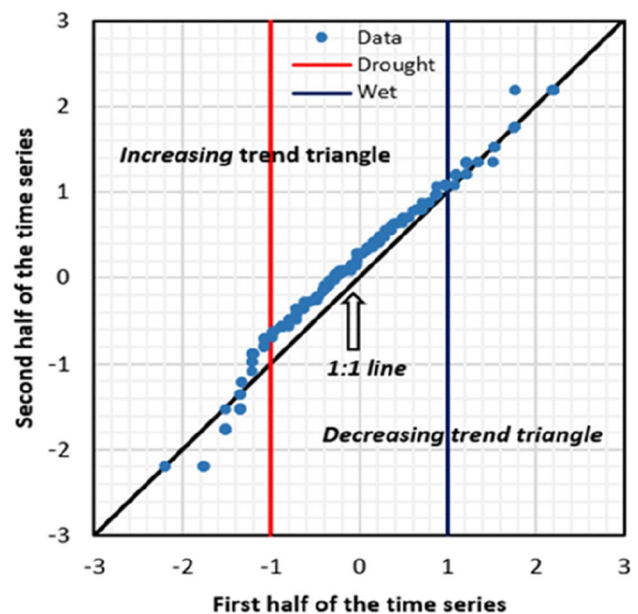


Fig. 2 Schematic example of the innovative trend analysis

3.3 Machine learning models

This study employed four ML models to predict SPI over various timeframes. Furthermore, Python programming language was used to implement ML models in the present study. To be more specific, the Scikit-learn library was utilized for implementing the MLR and KNN models, whereas the XGBoost library was used for applying XGBoost. Moreover, ANN was implemented using the Keras library with TensorFlow as backend.

Prior to training an ML model, the dataset can be subjected to important modifications, which may have a significant impact on the model performance (Bisong 2019). For instance, a scaler technique called MinMaxScaler from the Scikit-learn library was used in this study. It basically rescales each variable within a range between 0 and 1. Lastly, to train ML models, the data was randomly divided into train dataset (80% of the data) and test dataset (the rest of the data). In the following, each ML model is presented:

3.3.1 Multiple Linear Regression

MLR assumes a linear relationship between independent variables (input data denoted by x_i) and the dependent variable (output data denoted by y) (Choubin et al. 2016). Equation 4 is the general form of an MLR model:

$$y = c_0 + c_i x_i \text{ for } i = 1, 2, \dots, n \quad (4)$$

where c_0 is a constant value representing the point that the regression line intersects the y -axis. The slopes of the lines connecting the regression line to each data point are represented by the c_i values. Lastly, n is the number of independent variables used to predict the dependent.

3.3.2 Artificial Neural Network

ANN is a widely used ML that resembles human brain (Niazkar 2020). In this study, the activation function for the hidden layer(s) was *tanh*, while the output layer was *linear*. The number of input neurons in each ANN model ranged from 1 to 6, depending on lead times. Furthermore, this study employed one hidden layer. Following a suggestion from a previous study (Mishra and Desai 2006), the number of hidden neurons in ANN was set to $2n + 1$, where n represents the number of input neurons. In addition, the ANN were trained with 500 epochs, while an early stopping criterion was also employed; if the error does not decrease after 100 epochs, the algorithm terminates the training process, and the weights and biases corresponding to the minimum error are returned.

3.3.3 K Nearest Neighbors

KNN is a popular ML model that predicts test data points based on classifications of their K Nearest Neighbors, i.e. train data points. Three advantages of KNN are the capacity to test multiple combinations and eliminate unimportant combinations to avoid overfitting, effective with large datasets, and robust to noisy data (Fadaei-Kermani et al. 2017). Distance functions are used to compare test and training data points. Minkowski distance function is one of the common distance functions described in the following equation:

$$d = \left[\sum_{i=1}^n (tr_i - te_i)^p \right]^{\frac{1}{p}} \quad (5)$$

where tr and te represent the train and test data, respectively, and p is the power of Minkowski distance function. If $p = 1$, the distance function is called Manhattan, while if $p = 2$, it is called Euclidean. The latter function was particularly utilized in this study.

After calculating the distances, they are sorted, and the nearest neighbour is determined based on the minimum distance, i.e. the maximum similarity. KNN predicts output of a given data point by finding the K nearest data points based on the maximum similarity and calculating the weighted average target values of the data points. Selecting an adequate value for the number of nearest neighbours, i.e. K , is essential. A large value for K may result in including out-class data points, while a small one may not efficiently train the model and can be sensitive to noises. The optimal value of K can be determined through a cross-validation process. In this study, the number of neighbours was set to 5.

3.3.4 eXtreme Gradient Boosting for regression

XGBoost is a tree-based boosting ML method that is known for its performance, speed, scalability, and its unique features for efficiently handling large datasets (Chen and Guestrin 2016). Unlike traditional methods that average independent trees, such as Random Forest, XGBoost creates a series of consecutive decision trees by using prediction errors (residuals) of the previous trees (Kumar et al. 2023). This approach allows XGBoost to focus on data points with more uncertainty (Abedi et al. 2022). Finally, the weighted average of the results of the trees leads to a final output. XGBoost can be used for both regression (XGBR) and classification (XGBC) tasks (Piraei et al. 2023).

XGBR consists of various hyperparameters that reduce overfitting, prediction variability, and improve accuracy. While most hyperparameters are set to their default values, certain parameters are tuned for each model: (i) $n_estimators$ determines the maximum number of trees, (ii) $learning_rate$ controls the weight assigned to each tree to capture data

patterns, (iii) *reg_alpha* and (iv) *reg_lambda* control L1 and L2 weight regularization terms to prevent overfitting, respectively. Each hyperparameter was tuned specifically for each scenario, while the following intervals were considered for the aforementioned hyperparameters: (i) [100, 2000], (ii) [0.1, 1], and (iii and iv) [0, 2], respectively.

3.4 Performance metrics

To evaluate the performance of different ML models, four criteria were employed: (i) Root Mean Square Error (RMSE), (ii) Mean Absolute Errors (MAE), (iii) Nash–Sutcliffe efficiency (NSE), and (iv) Determination coefficient (R^2). The equations of these metrics are shown in the following (Niazkar and Zakwan 2023):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (7)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n \left(O_i - \frac{\sum_{i=1}^n O_i}{n}\right)^2} \quad (8)$$

$$R^2 = \left\{ \frac{\sum_{i=1}^n \left[\left(O_i - \frac{\sum_{i=1}^n O_i}{n}\right) \left(P_i - \frac{\sum_{i=1}^n P_i}{n}\right) \right]}{\sqrt{\sum_{i=1}^n \left(O_i - \frac{\sum_{i=1}^n O_i}{n}\right)^2 \sum_{i=1}^n \left(P_i - \frac{\sum_{i=1}^n P_i}{n}\right)^2}} \right\}^2 \quad (9)$$

In these equations, the symbols n , O , and P represent the total number of data points, observed, and estimated SPIs, respectively. Based on the definitions provided for each metric, an improvement in the accuracy of SPI estimations is associated with higher values of R^2 and NSE, as well as lower values of RMSE and MAE.

4 Results

4.1 Evaluation of the 6-, 9-, 12-, and 24-month drought trends

In this study, the ITA method was applied to the 6-, 9-, 12-, and 24-month SPI series to determine the possible meteorological drought trend in EBSB. The rainfall records of 8 stations in EBSB, which were determined to have sufficient record length and homogeneous structure

(Hirca et al. 2022), were used. In this study, two vertical bands were added in Fig. 2 to better understand the possible trends of dry and wet conditions: a red band corresponding to the drought limit ($\text{SPI} = -1$) and a blue band corresponding to the wet limit ($\text{SPI} = 1$). Normal conditions are represented by the region between the two bands, which makes it possible to show both low and high SPI trends with the ITA approach. Figures 3, 4, 5, and 6 show regional results of the ITA methodology applied to the 6-, 9-, 12-, and 24-month SPI series.

The ITA results obtained using 6-month SPI values are given in Fig. 3. As shown, the highest and lowest values (except for station 4) of the 6-month SPI results are trendless at all stations. Therefore, there is no significant change in extremely dry and extremely wet conditions. Furthermore, a decreasing trend was observed in normal conditions ($-1 < \text{SPI} < 0$) and dry conditions in the 6-month evaluation for station 2, which leads to more severe droughts and weaker wet periods. Moreover, the decreasing trend in drought indices in the dry period at station 3 revealed that the severity of droughts increased, while wet conditions did not show a clear trend. According to Fig. 3, the increasing trend of moderate wetness ($1.00 \leq \text{SPI} < 1.50$) in wet conditions in station 5 indicates that moderate wetness has increased. Although a decreasing trend is observed in station 6 in the dry period as SPI decreases, the situation is the opposite in the wet period, and it is observed that the SPI trend enhances with the increase of SPI. In station 7, a sharp decrease and a sharp increase are observed in the dry and wet season trends, respectively. As a result, it demonstrates that there are more severe droughts and heavier wet periods in station 7. Additionally, station 1 and station 8 illustrate very similar trends in normal and wet conditions. The increasing trend in both conditions indicates that more heavy wet periods can be experienced. In addition, the increasing trend in dry conditions in station 1 indicates a weaker dry period.

The ITA results obtained by the 9-month SPI values are given in Fig. 4. As shown for station 2, there is generally a decreasing trend as the SPI values decrease under dry ($\text{SPI} \leq -1$) and normal conditions. This indicates more severe dry periods. In station 3, there is a sharply decreasing trend in negative SPI values, especially for dry conditions, and a slightly increasing trend in wet periods. Unlike station 3, there was no clear trend in wet conditions in station 4. In station 5, an increasing trend was observed in wet conditions, which has similar characteristics as that of the 6-month SPI. In station 6, no trend was detected under normal conditions. However, although there is a slightly decreasing trend in the dry period, there is an increasing trend in wet conditions. In station 7, on the other hand, the increasing trend of positive SPI values indicates that more severe extreme wet conditions are dominant. In station 8 and station 1, there has

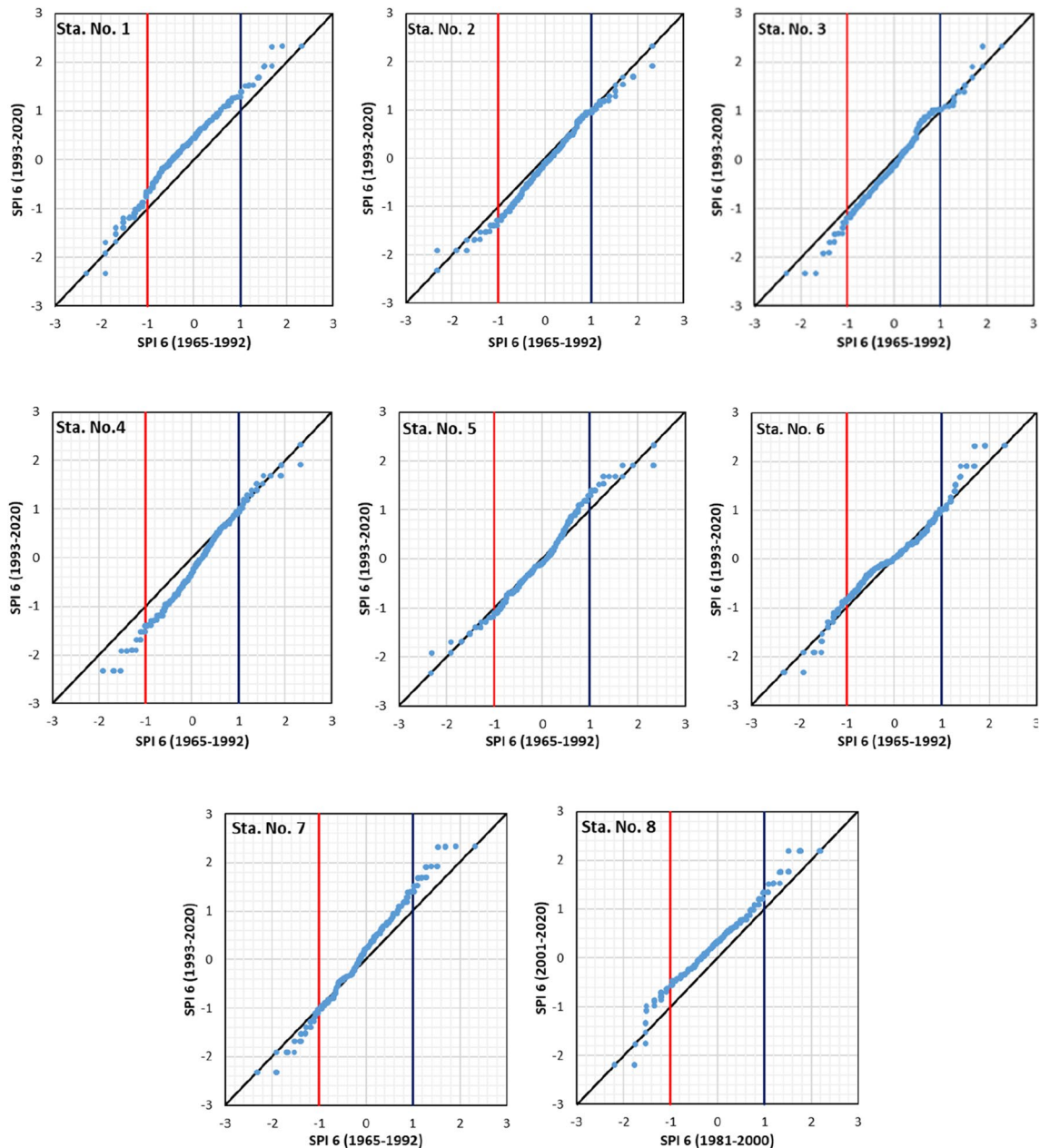


Fig. 3 Results of ITA applied to the 6-month SPI values

been a trend for all SPI values to increase, leading to milder droughts and more severe wet periods.

According to the 12-month SPI results given in Fig. 5, although there is no trend in the highest and lowest values of SPI in station 2, the decreasing trend in negative SPI values causes more severe droughts. In station 3, the zero-SPI value is the turning point of the trend (from decreasing to increasing). Accordingly, the increasing trend of positive SPI values and decreasing trend of negative SPI values reveal that the dry periods are drier and wet periods are wetter. However, when the trends in negative and positive SPI

values are examined, it is possible to say that droughts are more severe. The decreasing trend of negative SPI values in station 4 also reveals that dry periods are more severe, like that of station 3. Unlike the trend for station 3, a decreasing trend is observed in station 4 under normal conditions (in station 3, the zero-SPI value was the turning point of the trend). An increasing trend was observed in station 5 during the moderate wet period ($1.00 \leq \text{SPI} < 1.50$). Although a slightly increasing trend was observed in dry and wet periods in station 6, no significant trend was found in general. The sharp transitions in positive and negative SPI values in

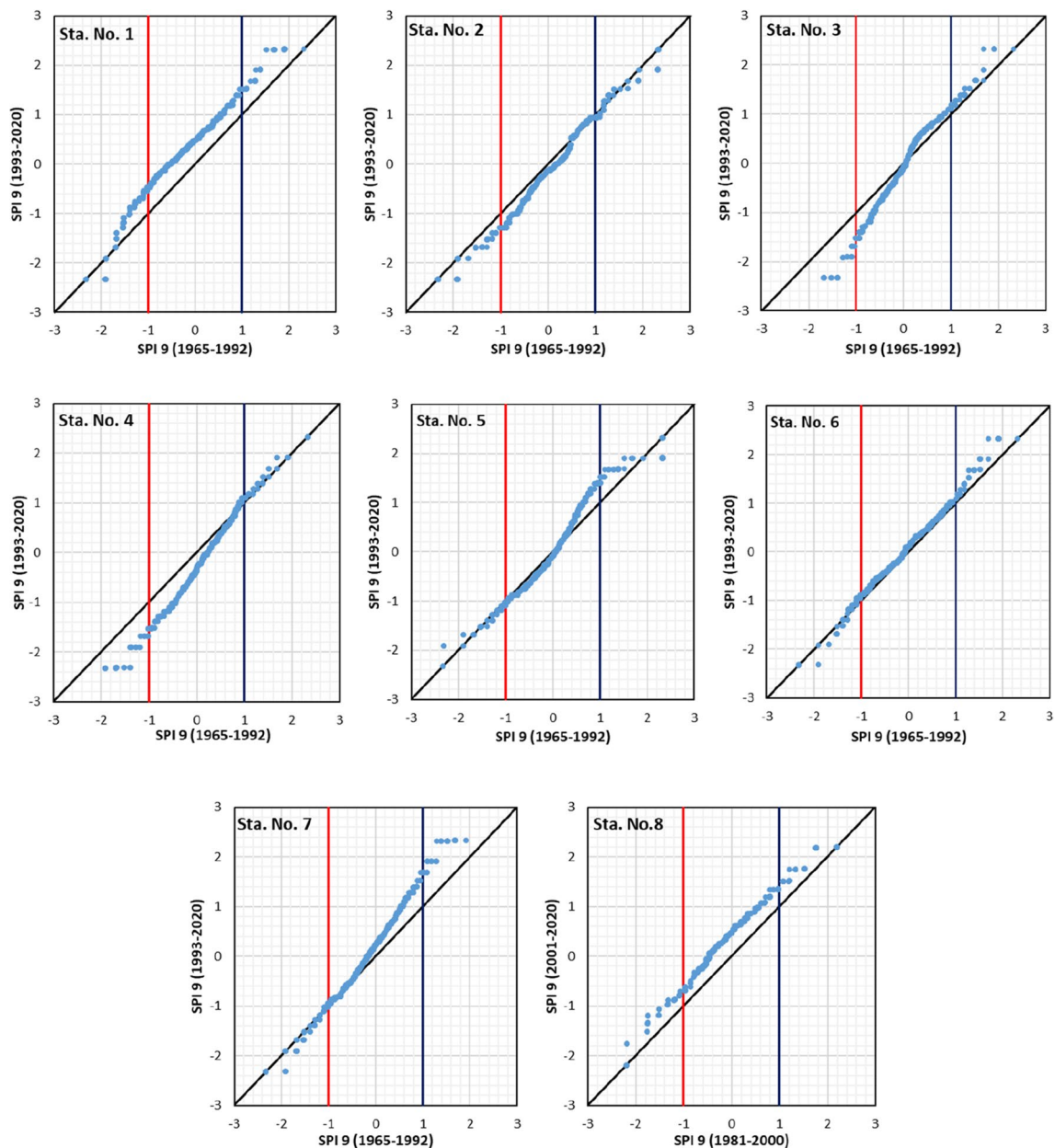


Fig. 4 Results of ITA applied to the 9-month SPI values

station 7 lead to more severe droughts and heavier wet periods. Furthermore, there has been a tendency for an increase in all SPI values in all conditions of station 8 and in station 1—generally—leading to weaker droughts and more heavy wet periods. In addition, the increasing trend in wet conditions in station 8 means that wet conditions are increasing. Based on the distribution of the lowest SPI values in the Cartesian coordinate system in Fig. 5, station 8 experienced more severe droughts during 1981–2000 compared to 2001–2020 because when the annual total rainfall averages in the two periods were compared, a difference of 81.89 mm was obtained.

According to the trend results of the 24-month SPI values, in station 2, there was a tendency to decrease SPI values in dry conditions, which led to more severe droughts. However, negative SPI values follow a monotonically decreasing trend, and the decrease is weaker during wet periods. Around the zero-SPI value in station 3, there are decreasing and increasing trends in dry and wet periods, respectively. This situation causes the dry periods to be more severe. However, there is a slightly increasing trend in wet periods. Although there is a slightly increasing trend in wet conditions in station 4, droughts have intensified in the region. Although there is no trend in the lowest

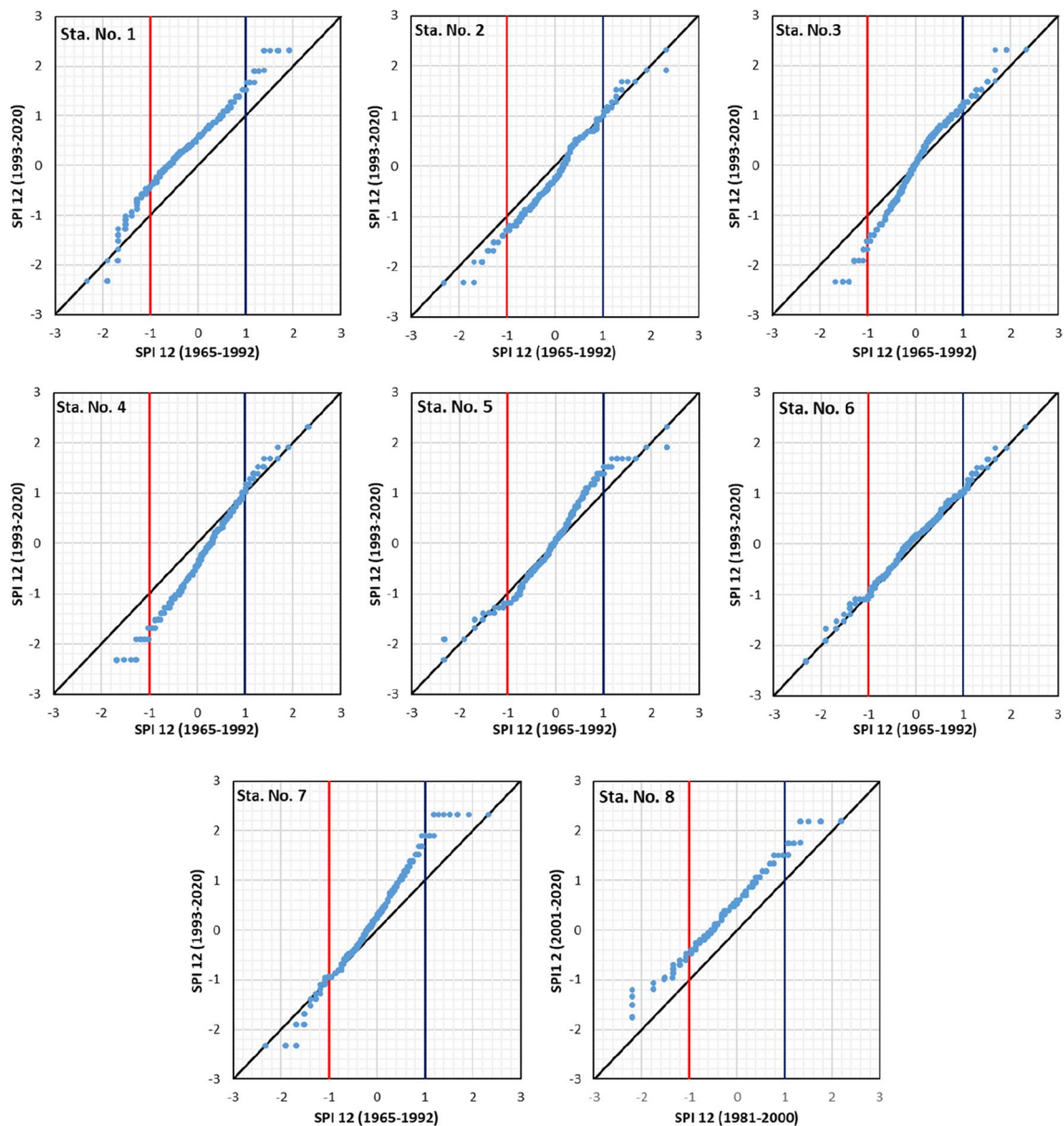


Fig. 5 Results of ITA applied to the 12-month SPI values

and highest SPI values in station 5, there is a decreasing trend in dry and wet periods. There has been a tendency for SPI values to increase in dry and wet conditions in station 6, resulting in weaker droughts and heavier wet periods. For station 7, a sharp increase trend in normal and wet conditions and a decreasing trend for dry conditions are remarkable, which causes more severe droughts and more severe wet periods. The decreasing trend, which started to be felt gradually in the 9- and 12-month SPI values in station 1, became stronger in the 24-month SPI values. Finally, a tendency to increase all SPI values in station 8 leads to weaker droughts and heavier wet periods.

4.2 Metric results for different SPIs

This study utilized 4 ML models to predict 6-, 9-, 12-, and 24-month SPI values for 8 stations considering 6 different lead times. Thus, 768 models were generated in total. To evaluate these models, Figs. 7, 8, 9, and 10 depict the results of 4 statistical metrics using heatmaps.

Figures 7, 8, 9, and 10 correspond to SPI values at 6, 9, 12, and 24 months, respectively. In Figs. 7, 8, 9, and 10, each row consists of four heatmaps that display the results for four metrics: RMSE, MAE, NSE, and R^2 . Likewise, each column represents one of the ML models: ANN, MLR, KNN, and

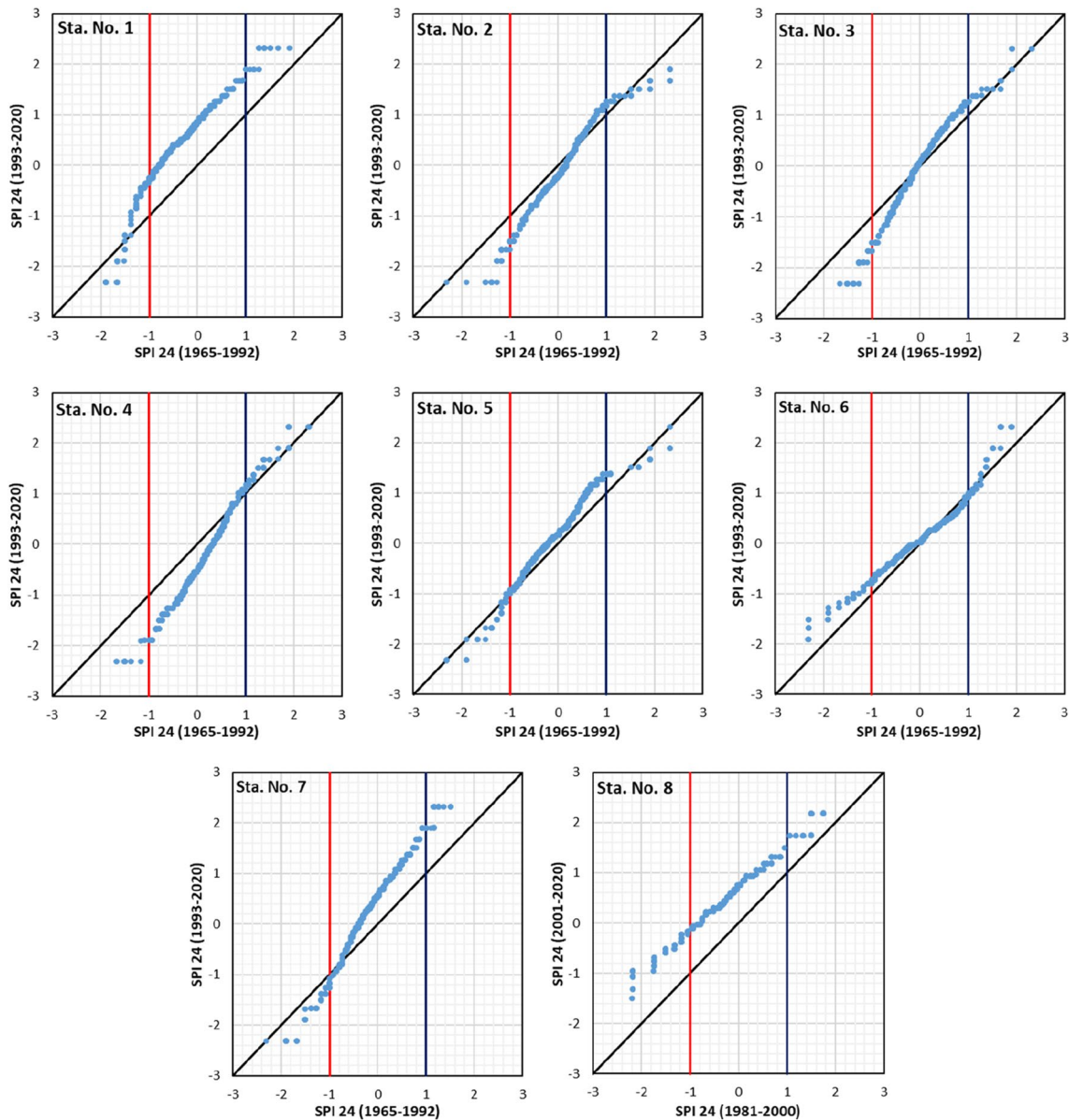


Fig. 6 Results of ITA applied to the 24-month SPI values

XGBR. The y-axis of each heatmap denotes different lead times, ranging from 1 to 6, while the x-axis represents the eight different stations. The heatmaps present a fixed colour range to demonstrate the performance of estimation models, with superior performance indicated by blue and inferior one by grey. According to Figs. 7, 8, 9, and 10, the results highlight a variation in performances for predicting SPI values. Generally, the ML models performed better for SPIs at 24 months, followed by 12 and 9 months, respectively, while they relatively demonstrate a poorer performance for the 6-month SPI. Finally, the metrics obtained for various scenarios are presented in a table in supplementary materials section.

5 Discussion

Analysis of drought indices has an important place in studies on climate change and extreme events. Within the scope of this study, SPI indexes were calculated using the monthly total rainfall data of 8 stations in EBSB. In this context, drought indices were calculated at various time steps to compare and evaluate medium (6- and 9-month SPI) and long (12- and 24-month SPI) droughts. The temporal trends of SPI values were examined by ITA, while ML-based estimation models were developed to forecast SPI indices. In the following, the obtained findings are discussed.

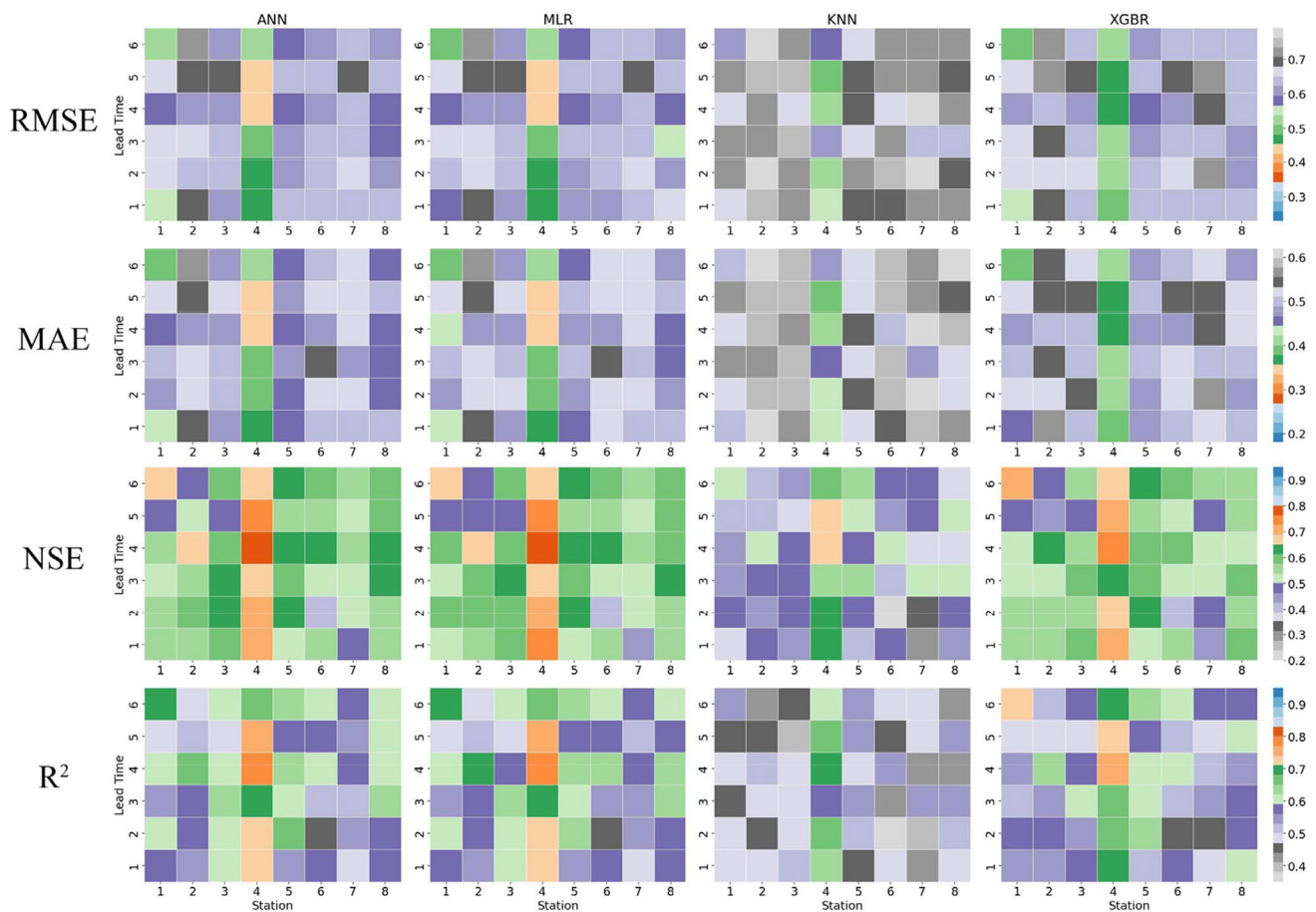


Fig. 7 Heatmap of the metric results for all the models utilized in this study for SPI6

5.1 Discussion of ITA results

It is crucial to study and analyze the drought to take necessary precautions against drought. In this study, ITA, which does not require any assumption, was used to examine the changes in drought indices. In this regard, correct interpretation of trends obtained by ITA results is important to underline drought conditions in the study area. In essence, the decreasing trend in negative SPI values means that droughts are intensified (increasing trend occurs in alleviating droughts), while the decreasing trend in positive SPI values implies weaker rainfall (increasing trend occurs in more severe wet conditions). Thanks to the bands added to the Cartesian coordinate system, it allows the determination of trends in dry, normal, and wet conditions. Therefore, it is a very useful method as it can be used as a preliminary assessment for more comprehensive climate studies.

Generally, 6- and 9-month SPI values can be used in the evaluation of agricultural drought that occurs after meteorological drought. It is expressed as a situation where there is not enough moisture in the root zone of the plant for its growth and development. It is vital to investigate agricultural

drought because it will cause a decrease in the amount of product or a change in growth. In the evaluation made within the scope of this study, the changes in the 6- and 9-month SPI values in most stations support each other.

Twelve- and 24-month SPI values can be used in the evaluation of hydrological drought, which is caused by the decrease in surface and groundwater because of a long-term lack of rainfall. The fact that station 8 has an increasing trend in all conditions for both periods reveals that the region has a milder drought period and a heavier wet period.

5.2 Discussion of ML-based SPI estimations

This study utilized 4 metrics (i.e. RMSE, MAE, NSE, and R^2) to evaluate the performance of different ML models for estimating SPI. Since Figs. 7, 8, 9, and 10 have the same range of colours for metrics, the results of the ML models can be compared for different stations and SPI values. As shown, it can be noted that as the number of SPI months increases, the accuracy of the estimation models improves, which is consistent with the trend analysis results. In the

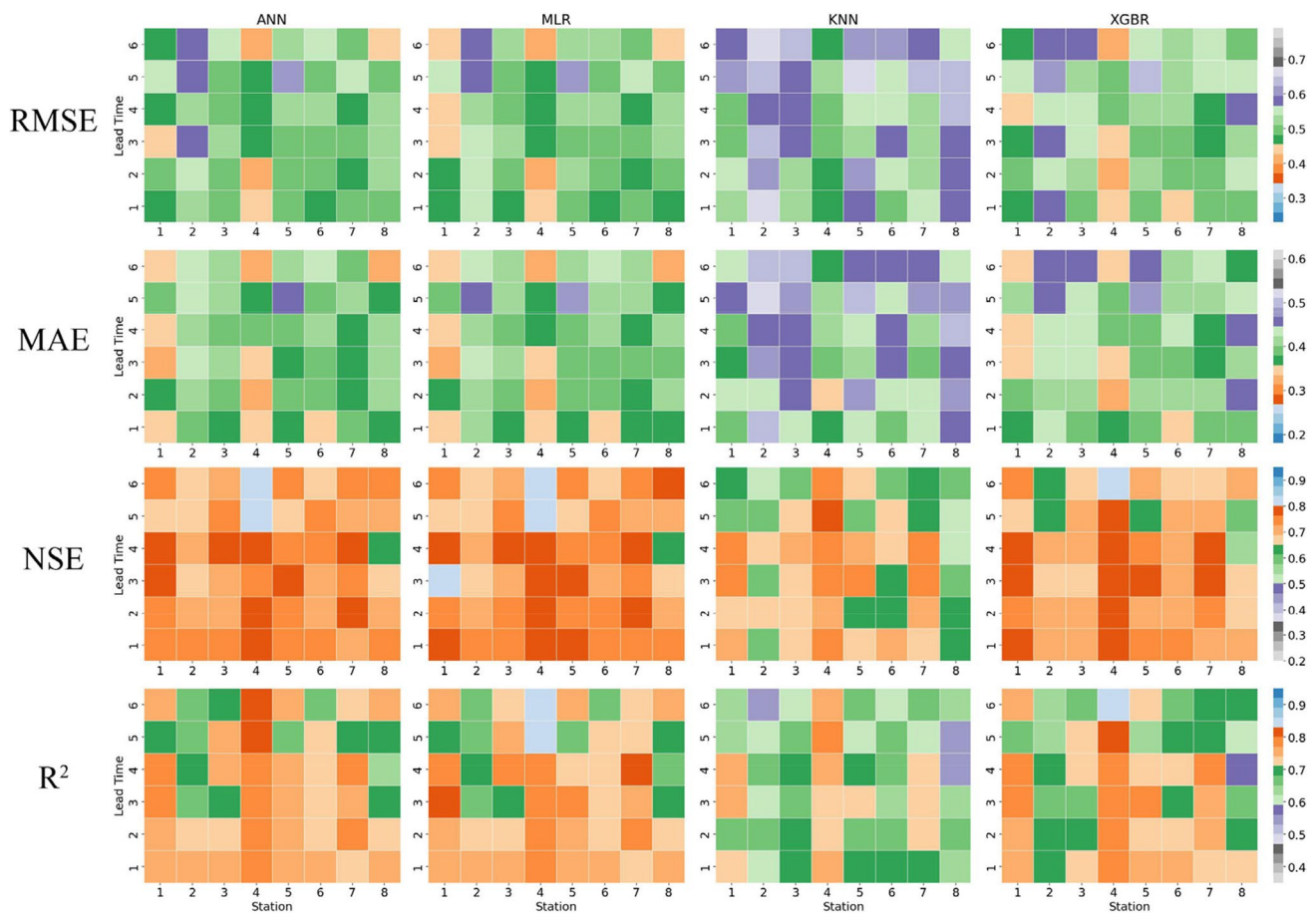


Fig. 8 Heatmap of the metric results for all the models utilized in this study for SPI9

following, the performance of ML models for predicting SPI values is discussed:

5.2.1 6-month SPI

According to Fig. 7, a total number of 192 models were generated from 4 ML model (i.e. ANN, MLR, KNN, XGBR) for the 6-month SPI. Based on various scenarios available, several outcomes can be obtained on a station-by-station basis:

For station 1, the XGBR model with a lead time of 6 exhibited the best performance based on RMSE (0.489), NSE (0.701), and R² (0.710), as shown in Fig. 7. However, considering the MAE criterion, the MLR model with a lead time of 6 outperformed other models (MAE = 0.385). These results indicate that a lead time of 6 yields the best performance for station 1.

For station 2, the ANN model with a lead time of 4 showed the best performance in terms of RMSE (0.602) and NSE (0.670), while the MLR model with a lead time of 4 performed the best in terms of MAE (0.477) and R² (0.685). Thus, a lead time of 4 is optimal for station 2.

In the case of station 3, the ANN model with a lead time of 4 achieved the lowest RMSE (0.604), while the MLR model with a lead time of 4 had the lowest MAE (0.468). However, for NSE and R², the MLR model with a lead time of 3 demonstrated superior performance with values of 0.619 and 0.624, respectively. These findings suggest that utilizing lead times of 3 or 4 may yield better performance for station 3.

The results of SPI estimations for station 4 exhibited the best performance by the ANN model with a lead time of 4 based on all metrics (RMSE = 0.439, MAE = 0.338, NSE = 0.773, and R² = 0.773). Hence, a lead time of 4 is optimal for station 4.

Similarly, the ANN model with a lead time of 4 showed the best performance for station 5 in terms of RMSE (0.576), while the ANN model with a lead time of 6 achieved the lowest MAE (0.447). Additionally, the ANN model with a lead time of 2 performed the best in terms of NSE (0.649) and R² (0.651). Although the ANN model outperformed other ML models for station 5, no definitive conclusion can be drawn regarding the choice of lead time.

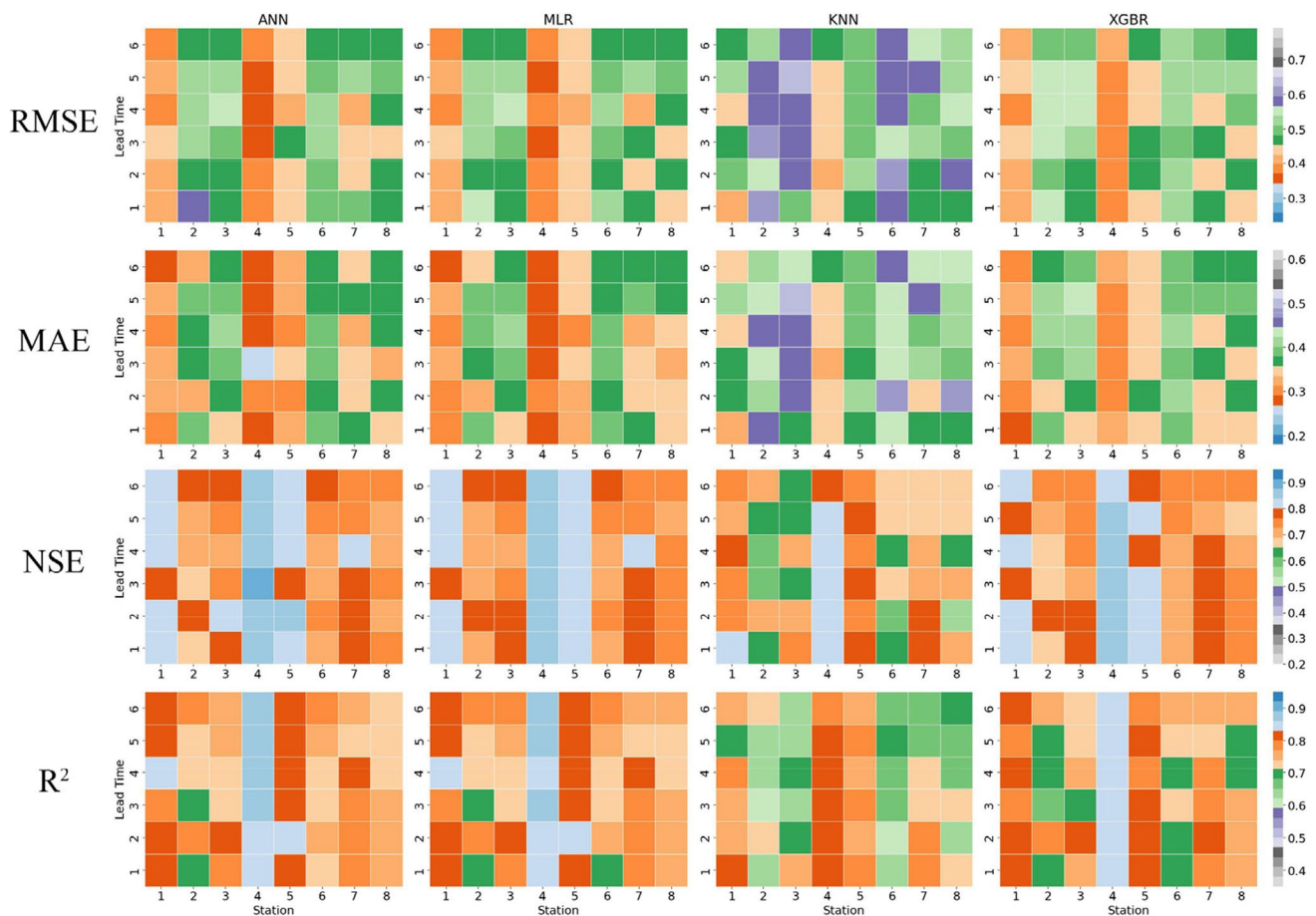


Fig. 9 Heatmap of the metric results for all the models utilized in this study for SPI12

For station 6, the MLR model with a lead time of 4 exhibited the best performance based on RMSE (0.605), MAE (0.480), NSE (0.621), and R^2 (0.621). Moreover, the ANN model with a lead time of 4 exhibited the second-best performance based on RMSE (0.609), MAE (0.484), NSE (0.616), and R^2 (0.617). Hence, a lead time of 4 is optimal for station 6.

The best estimations for station 7 were obtained by the ANN model with a lead time of 1 in terms of RMSE (0.624), while the ANN model with a lead time of 3 achieved the lowest MAE (0.484). Moreover, the ANN model with a lead time of 6 demonstrated the best performance for NSE (0.566) and R^2 (0.574). Nevertheless, like station 5, no conclusion can be made regarding the choice of lead time for station 7.

For station 8, the MLR model with a lead time of 3 yielded the best performance based on RMSE (0.566), MAE (0.453), NSE (0.626), and R^2 (0.630). On the other hand, the KNN model with a lead time of 6 yielded the worst performance based on RMSE (0.733), MAE (0.600), NSE (0.376), and R^2 (0.433). Overall, a lead time of 3 is an optimal choice for station 8.

Generally, the ML models with a lead time of 4 demonstrated superior performance, suggesting it to be a suitable choice for the 6-month SPI. The ANN model overall outperformed other models for the test data. Although the XGBR model performed slightly worse than ANN for the test data, it demonstrated higher accuracy for the train data. This indicates that choosing the XGBR model could be a viable alternative to ANN and MLR because it is less time-consuming and computationally costly. Furthermore, the KNN model demonstrated high accuracy for the train data. However, despite its acceptable performance on the test data, it was the weakest model in the comparative analysis. Due to the risk of overfitting associated with KNN, it is recommended to use it cautiously. Lastly, Fig. 7 illustrates that most heatmaps have a purple colour, indicating an average performance of the ML models in the 6-month SPI compared to other SPIs.

5.2.2 9-month SPI

Regarding 9-month SPI, 192 models were generated and are interpreted based on each rainfall station:

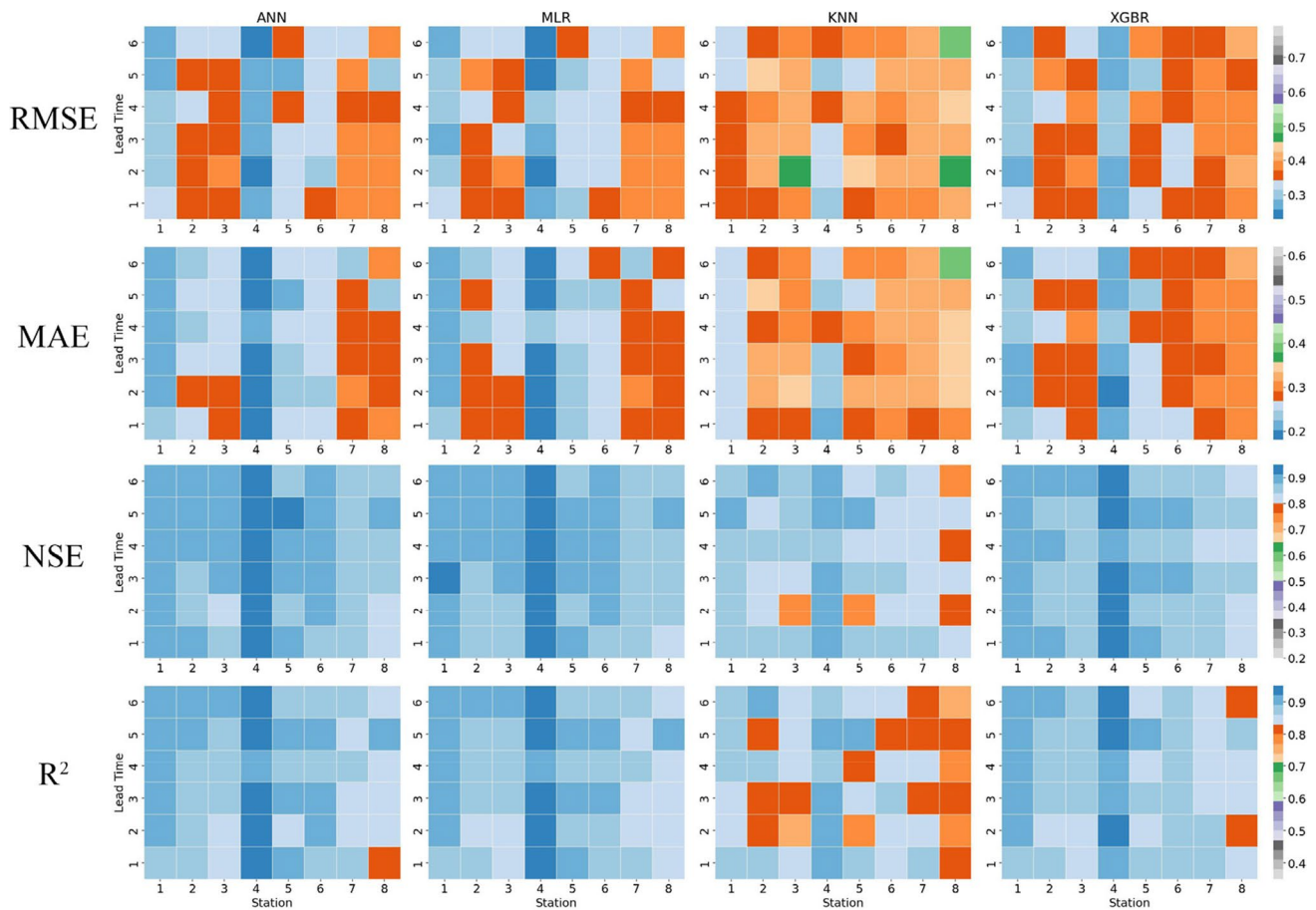


Fig. 10 Heatmap of the metric results for all the models utilized in this study for SPI24

For station 1, the MLR model with a lead time of 3 demonstrated the best performance across RMSE (0.430), MAE (0.327), NSE (0.802), and R^2 (0.803). Hence, a lead time of 3 is optimal for station 1.

For station 2, the ANN model with a lead time of 1 showed the best performance regarding the RMSE (0.528), MAE (0.397), and NSE (0.739), while the MLR model with a lead time of 1 outperformed other ML models in terms of R^2 (0.746). These findings suggest that a lead time of 1 yields the best performance for station 2.

For station 3, the MLR model with a lead time of 1 achieved the lowest RMSE (0.481), while the ANN model with a lead time of 1 achieved the best MAE (0.367). Additionally, the MLR model with a lead time of 4 yielded the best values of NSE (0.776) and R^2 (0.779). Thus, utilizing lead times of 1 or 4 can result in improved performance for station 3.

In the case of station 4, the MLR model with a lead time of 6 obtained the lowest RMSE (0.401), while the ANN model with a lead time of 2 demonstrated a superior performance in terms of MAE (0.315). Moreover, the former with a lead time of 6 exhibited the best performance based

on NSE (0.830) and R^2 (0.833). These results indicate that employing lead times of 2 or 6 can lead to better performance for station 4.

For station 5, the XGBR model with a lead time of 3 outperformed other ML models in terms of RMSE, NSE, and R^2 , with values of 0.491, 0.769, and 0.773, respectively. The ANN model with a lead time of 1 achieved the lowest MAE (0.371). These results suggest that considering lead times of 1 or 3 may result in a better performance for station 5.

For station 6, the XGBR model with a lead time of 1 delivered the best performance based on all metrics. To be more specific, it obtained RMSE=0.451, MAE=0.342, NSE=0.762, and R^2 =0.763. Therefore, a lead time of 1 is optimal for station 6.

The SPI data of station 7 was forecasted the best by the MLR model with a lead time of 4 because it achieved the lowest RMSE (0.459) and MAE (0.360), and highest NSE (0.800) and R^2 (0.801). Hence, a lead time of 4 is optimal for station 7.

Similarly, for station 8, the MLR model with a lead time of 6 showcased the best performance in terms of RMSE, NSE, and R^2 , with values of 0.430, 0.763, and 0.769,

respectively. However, the ANN model with a lead time of 6 exhibited the lowest MAE (0.333) for the case. These results indicate that a lead time of 6 yields the best performance for station 8.

Overall, the interpretation of ML model performances was consistent with the 6-month SPI, with the ANN, MLR, and XGBR models demonstrating good performances and the KNN model demonstrating acceptable one. Furthermore, despite significant variations in the outcomes, the application of lead time 1 yielded highly satisfactory results within this study. Furthermore, Fig. 8 reveals that the predominant colour observed in the heatmaps for the RMSE and MAE fell within the green spectrum, indicating an average performance of the ML models compared to other SPIs. Conversely, for the NSE and R^2 metrics, the primary colour exhibited in the heatmaps is in the colour range of red, suggesting a notably higher performance than the average and satisfactory results of the ML models.

5.2.3 12-month SPI

Figure 9 illustrates 192 scenarios for 12-month SPI, which are discussed through a station-based order in the following:

For station 1, the MLR model with a lead time of 4 demonstrated the best performance in terms of RMSE (0.378) and R^2 (0.836). Additionally, the ANN model with a lead time of 6 exhibited the best performance based on MAE (0.286), while the MLR model with a lead time of 2 yielded the highest NSE (0.830). Nevertheless, no definitive conclusion can be drawn regarding the optimal lead time for this station.

For station 2, the MLR model with a lead time of 2 achieved the best performance in terms of RMSE (0.459), NSE (0.796), and R^2 (0.798). On the other hand, the ANN model with a lead time of 2 demonstrated the best performance in terms of MAE (0.314). Therefore, a lead time of 2 is recommended for station 2.

For station 3, the ANN model with a lead time of 2 reached the best performance based on RMSE (0.455), NSE (0.808), and R^2 (0.811), while the MLR model with a lead time of 1 achieved the lowest MAE (0.344). These results suggest that using lead times of 1 or 2 can improve the estimation of SPI values for station 3.

For station 4, the ANN model with a lead time of 3 yielded the best performance based on RMSE (0.353), MAE (0.266), and NSE (0.875), while the MLR model with a lead time of 3 yielded the best performance in terms of R^2 (0.877). Therefore, a lead time of 3 is optimal for station 4.

For station 5, the best performance in terms of RMSE (0.420) and MAE (0.308) was achieved by the ANN model with a lead time of 4. Additionally, the ANN model with a lead time of 2 achieved the highest NSE (0.839) and R^2

(0.841). These findings suggest that lead times of 2 or 4 may lead to improved performance for station 5.

In the case of station 6, the MLR model with a lead time of 6 yielded the best performance based on RMSE (0.472), NSE (0.784), and R^2 (0.784), while the ANN model with a lead time of 2 achieved the lowest MAE (0.360). Thus, utilizing lead times of 2 or 6 may improve performance for station 6.

The results obtained for station 7 showed that the ANN model with a lead time of 4 achieved the best metrics: RMSE (0.412), MAE (0.323), NSE (0.812), and R^2 (0.816). Therefore, a lead time of 4 is optimal for station 7.

For station 8, the MLR model with a lead time of 3 demonstrated the best performance in terms of RMSE (0.436), NSE (0.761), and R^2 (0.765), while the ANN model with a lead time of 3 achieved the lowest MAE (0.329). Hence, a lead time of 3 is recommended for station 8.

The performance interpretation of the ML models was consistent with 6- and 9-month SPIs, with the ANN, MLR, and XGBR models displaying good performances, while the KNN model showed an acceptable one. Moreover, lead times of 2 to 4 yielded promising results for the 12-month SPI, considering the variability of the outcomes. When analyzing Fig. 9, the heatmaps predominantly showed green and red colour ranges for RMSE and MAE criteria, indicating an above-average overall performance. Additionally, the colour range of red dominated the heatmaps for NSE and R^2 criteria, which indicates satisfactory precision.

5.2.4 24-month SPI

As depicted in Fig. 10, there are 192 various scenarios for 24-month SPI. The estimated SPI values are discussed for different stations separately in the following:

For station 1, the MLR model with a lead time of 6 exhibited a superior performance in terms of RMSE (0.274). Likewise, the ANN model with a lead time of 6 demonstrated the best performance based on MAE (0.203), while the MLR model with a lead time of 3 outperformed other ML models in terms of NSE (0.913) and R^2 (0.914). These findings suggest that employing lead times of 3 or 6 can enhance the SPI predictions for station 1.

Regarding station 2, the best performance belongs to the ANN model with a lead time of 6 as the ANN-based estimations obtained the lowest RMSE (0.327), and highest NSE (0.900) and R^2 (0.901). Also, the ANN model with a lead time of 4 yielded the best MAE (0.239). These results indicate that lead times of 4 or 6 are conducive to improve SPI forecasting for station 2.

The optimal lead time for station 3 was found to be 6 with the ANN model as it achieved the lowest RMSE (0.323) and MAE (0.246), and highest NSE (0.892) and R^2 (0.893). On the other hand, the KNN model with a lead time of 2

presented the weakest performance based on RMSE (0.458), MAE (0.353), NSE (0.760), and R^2 (0.766).

Similarly, for station 4, the ANN model with a lead time of 2 presented a superior performance based on RMSE (0.236), MAE (0.185), NSE (0.947), and R^2 (0.947). Hence, a lead time of 2 is recommended for station 4. Furthermore, the KNN model with a lead time of 4 performed the worst based on RMSE (0.349), MAE (0.274), NSE (0.873), and R^2 (0.874).

Station 5 exhibited the best performance with the ANN model and a lead time of 5, which is characterized by the lowest RMSE (0.281) and MAE (0.211), and highest NSE (0.917) and R^2 (0.918). Therefore, a lead time of 5 is optimal for station 5.

Regarding station 6, the ANN model with a lead time of 2 demonstrated the lowest RMSE (0.313), while the MLR model with a lead time of 5 showcased the best MAE (0.241). Moreover, the MLR model with a lead time of 3 achieved the highest NSE (0.895) and R^2 (0.895). No definitive conclusion can be drawn regarding the choice of lead time for station 6.

For station 7, the MLR model with a lead time of 6 delivered the lowest RMSE (0.329) and MAE (0.245), while the XGBR model with a lead time of 1 outperformed other ML models in terms of NSE (0.873) and R^2 (0.876). Thus, using lead times of 1 or 6 can lead to accuracy-improved predictions for station 7.

The best performance for station 8 was observed with the ANN model and a lead time of 5, as indicated by the RMSE (0.314), MAE (0.241), NSE (0.903), and R^2 (0.904). Therefore, a lead time of 5 is recommended for station 8.

Like previous SPIs, the ANN, MLR, and XGBR models displayed commendable performances, while the KNN model exhibited an acceptable performance. Additionally, the ML models with a lead time of 6 generally outperformed others, making it a suitable choice for estimating 24-month SPI values. Analyzing Fig. 10 indicated that the predominant colours in most heatmaps were in the range of blue, which implies the strong performance of the ML models for estimating 24-month SPI. Thus, compared to other SPIs, the ML models exhibited a significantly improved performance for this SPI.

Generally, the results of NSE and R^2 in all 768 cases were very similar, indicating that all ML models were trained properly. Moreover, it is worth noting that almost similar metrics results were obtained across the ML models used in this study. It suggests that the ML models have similar predictive accuracy and are performing acceptably well in terms of minimizing the overall prediction errors. In such scenario, the best ML model for estimating SPI values depends on other factors, like computational efficiency, interpretability, ease of implementation, or specific requirements of the problem at hand. Moreover, the best ML model can be different

for each station, lead time, or return period of calculating drought index.

It is recommended that drought is not only dependent on the changes in rainfall but also examining the changes in different drought indices using hydroclimatic variables, such as temperature and evapotranspiration, in terms of more accurate and realistic planning of water resources in the region. Therefore, further studies on other drought indices, like SPEI and STD, are suggested.

5.2.5 Related work and future works

In this study, various models including ANN, MLR, KNN, and XGBR were employed to predict 6-, 9-, 12-, and 24-month SPIs with lead times ranging from 1 to 6. Table 2 presents a summary for an overall assessment of this study compared to previous studies in the literature. As shown, performances of the models used in this study demonstrated an enhancement in the accuracy of SPI predictions as the number of SPI months increased. Specifically, they exhibited their strongest performance in estimating the 24-month SPIs, while their weakest performance was observed in estimating the 6-month SPIs. This finding aligns with prior studies. For instance, Belayneh and Adamowski (2012) compared the performance of various ML models for forecasting 3- and 12-month SPIs for the Awash River Basin of Ethiopia. Their results indicated that all ML models yielded a higher accuracy for 12-month SPI compared to that of 3-month SPI.

The optimal performing model in this study varies depending on the station being analyzed, the specific SPI, and lead time considered. As a result, it is challenging to determine definitively which model performed the best compared to others. However, the XGBR, ANN, and MLR models demonstrated robust accuracy in predicting SPIs. This outcome is consistent with previous research. Most previous studies used ANN or its variants/hybrids and concluded that it demonstrates a robust performance (Belayneh and Adamowski 2012; Belayneh et al. 2014, 2016; Maca and Pech 2016; Khan et al. 2020; Ham et al. 2023). Moreover, Malik et al. (2021) predicted 1-, 3-, 6-, 9-, 12-, and 24-month SPIs for the Garhwal region of Uttarakhand State in India, employing Multi-Layer Perceptron Neural Network (MLPNN), Co-Active Neuro-Fuzzy Inference System (CANFIS), and MLR. They concluded that CANFIS is the optimal choice for majority of stations and SPIs, while the MLR model was the best choice for the rest.

According to the results, XGBR performed as well as ANN and MLR for the test dataset, while it yielded better performances for the training dataset. Additionally, the KNN model exhibited robust performances for the training dataset. However, its performance was poorer for the test dataset when it was compared to other ML models. Thus, it is the least favourable choice among models considered in

Table 2 Overview of ML models and lead times conducted in previous studies and this study

Study	Stations	SPI months	Lead times	ML models used for comparison	Best resulting model	Best number of lead times
Belayneh and Adamowski (2012)	3	3, 12	1, 6	ANN, SVR, and WN	WN	1
Belayneh et al. (2014)	12	12, 24	6, 12	ARIMA, ANN, SVR, WA-ANN, and WA-SVR	WA-ANN	6
Hosseini-Moghari and Araghinejad (2015)	1	3, 6, 9, 12, 24	1–6	RMSMLP, DMSMLP, RMSRBF, DMSRBF, RMSGRNN, and DMS-GRNN	RBF and GRNN, with recursive models (RM) performing better for SPI with smaller time scales, while direct models (DM) performed better for larger ones	1
Belayneh et al. (2016)	3	3, 12, 24	1–9	BANN, BSVR, WA-ANN, WA-SVR, BS-ANN, BS-SVR, WBS-ANN, and WBS-SVR	WBS-ANN and WBS-SVR	Different lead times based on trial-and-error
Maca and Pech (2016)	2	1	3, 6, 9	sANN and hANN	hANN	9
El Ibrahim and Baali (2018)	2	3, 12	1–3	ANFIS, ANN-MLP, and SVR	ANFIS	1
Khan et al. (2020)	3	1	3	WN, ARIMA, ANN, and W-2A	W-2A	3 (only choice)
Adikari et al. (2021)	13	1	24	CNN, LSTM, and WANFIS	WANFIS	24 (only choice)
Malik et al. (2021)	7	1, 3, 6, 9, 12, 24	1–12	ANN-MLP, CANFIS, and MLR	CANFIS and MLR	Different sets of lead times based on the autocorrelation function (ACF) and partial-ACF (PACF) analysis
Taylan et al. (2021)	3	3, 6, 9, 12	Not mentioned	ANFIS, SVM, ANN, WA-ANFIS, WA-ANN, and WA-SVR	WA-ANFIS for 12-month SPI and WA-SVR for other SPIs	Not mentioned
Pande et al. (2023)	3	3, 6, 12	1–15	AR, RSS, MSP, and BT	M5P and BT	Different sets of lead times based on the best subset regression analysis
Elbeltagi et al. (2023a)	1	3, 6, 12	1–10	RSS, RSS-M5P, RSS-RF, and RSS-RT	RSS-M5P	Different sets of lead times based on the best subset regression and sensitivity analysis
Elbeltagi et al. (2023b)	2	6, 12	1–10	RF, RT, and GPR-PUK kernel	RF	Different sets of lead times based on the best subset regression and sensitivity analysis
Ham et al. (2023)	6	6, 12	1–16 for SPI-6 and 1–34 for SPI-12	WLSTMN, WA-ANN, and WA-SVR	WLSTMN	Different lead times based on ACF analysis
This study	8	6, 9, 12, 24	1–6	ANN, MLR, KNN, XGBR	ANN, MLR, and XGBR	Different lead times for each station and months of SPI

Abbreviations: SVR, Support Vector Regression; WN, wavelet neural networks; ARIMA, Autoregressive Integrated Moving Average; WA-ANN, ANN coupled with wavelet transform; WA-SVR, SVR coupled with wavelet transform; RMSMLP, Recursive Multi-Step Multi-Layer Perceptron; DMSMLP, Direct Multi-Step Multi-Layer Perceptron; RMSRBF, Recursive Multi-Step Radial Basis Function; DMSRBF, Direct Multi-Step Radial Basis Function; RMSGRNN, Recursive Multi-Step Generalized Regression Neural Network; DMSGRNN, Direct Multi-Step Generalized Regression Neural Network; BANN, bootstrap ANN; BSVR, bootstrap SVR; BS-ANN, boosting ensemble with ANN; BS-SVR, boosting ensemble with SVR; WBS-ANN, wavelet boosting with ANN; WBS-SVR, wavelet boosting with SVR; sANN, feedforward multilayer perceptron; hANN, integrated neural network; ANFIS, neuro-fuzzy adaptive inference systems; ANN-MLP, Artificial Neural Network of Multi-Layered Perceptron; W-2A, hybrid ensemble of WN, ANN, and ARIMA; CNN, Convolutional Neural Network; LSTM, Long Short-Term Memory network; WANFIS, wavelet decomposition functions combined with ANFIS; CANFIS, Co-Active Neuro-Fuzzy Inference System; WA-ANFIS, ANFIS coupled with wavelet transform; AR, Additive Regression; RSS, Random Subspace, MSP, M5Pruned; BT, Bagging Tree; RF, Random Forest; RT, Random Tree; RSS-M5P, Hybridization of RSS with M5P; RSS-RF, Hybridization of RSS with RF; RSS-RT, Hybridization of RSS with RT; GPR-PUK kernel, Gaussian process regression; WLSTMN, Wavelet Long Short-Term Memory Network

this study. According to the literature, Fadaei-Kermani et al. (2017) utilized KNN to predict SPI based on rainfall data of Kerman, Iran, and highlighted that it predicted reasonable SPI in the study region. Future studies can explore applications of other ML models.

The results indicated that an optimal lead time considered for predicting SPI can vary for each SPI and station. Previous studies revealed a lack of consistency regarding the initial selection of lead times, with varying approaches employed by different researchers, as shown in Table 2. While certain studies have demonstrated that the utilization of the minimum lead time (typically 1) yields optimal outcomes (Belayneh and Adamowski 2012; Belayneh et al. 2014; Hosseini-Moghari and Araghinejad 2015; El Ibrahim and Baali 2018), other studies recommend using higher lead times for the SPI forecast (Maca and Pech 2016). Furthermore, diverse analyses have been employed in previous studies to determine the most favourable combination of preceding lag times (Malik et al. 2021; Elbeltagi et al. 2023a, 2023b; Pande et al. 2023; Ham et al. 2023). Overall, it is suggested that the selection of lead times remains an area warranting additional exploration and should be considered in future research endeavors.

6 Conclusion

In the context of sustainable water resource management, it is crucial to identify instances of drought and their trends. To conduct a comprehensive drought analysis in this study, we calculated SPI values for various time scales (6, 9, 12, and 24 months) using monthly rainfall data from eight monitoring stations in the EBSB region spanning the period from 1965 to 2020. The drought analysis was conducted using a trend analysis method and four ML models. The outcome of each application is presented in the following:

In the 6-month time scale, an increasing trend was observed in all conditions at station 1 located in the northeast of the basin. This situation reveals that weaker drought and heavier rainfall are experienced. In station 4 located in the south of the basin, it was observed that rainfall decreased, and drought events intensified with a decreasing trend in moderate and dry conditions in this time scale.

According to the 9-month SPI evaluations, the increasing trend in station 6, located in the west of the basin, in the wet period indicates that rainfall occurs more heavily, while the relatively decreasing trend in dry conditions reveals that more severe droughts are experienced. Evaluations of 6- and 9-month SPI values revealed that drought events intensified at stations 3 and 4. Since this means that the plant and soil cannot meet the water at the time they demand, decreases in the quality of plant production may occur. For 12-month SPI, an increasing trend was observed

in all conditions at station 8. Therefore, it was determined that rainfall increased, and milder droughts were experienced. In the 24-month SPI values, the decreasing trend in dry conditions, except for stations 1, 6, and 8, is an indication of the intensification of droughts. This causes prolonged rainfall scarcity to manifest in components of the hydrological system, such as source levels, runoff, and soil moisture. Furthermore, it was determined that drought analyses performed using SPI at 6-, 9-, 12-, and 24-month time scales generally gave compatible results. Finally, among these stations, stations 1, 5, and 7 stood out for their highest average annual rainfall during the study period. These three stations represent a rare occurrence in Türkiye, where annual rainfall exceeds 2000 mm. Furthermore, this study utilized four ML models to predict 6-, 9-, 12-, and 24-month SPI values for 8 stations in the Eastern Black Sea Basin in Türkiye. The results indicated that the performances of the ML models improve as the number of SPI months increases, with the more robust performance observed in estimating the 24-month SPIs. Furthermore, the ML model with the optimal performance varied depending on the station and specific SPI considered, which makes it challenging to recommend a definitive best model for all cases. Nevertheless, the XGBR, ANN, and MLR models consistently demonstrated robust accuracy in predicting SPIs. For each SPI, 8 rainfall stations with 6 different lead times were considered, which leads to 48 cases. For the 6-month SPI, ANN performed more robust based on RMSE in 21 out of 48 scenarios, while MLR and the XGBR performed better in 20 and 6 scenarios, respectively. Regarding the 9-month SPI, MLR outperformed others in 37 out of the 48 scenarios, while XGBR and ANN reach the best SPI predictions in 6 and 5 scenarios, respectively. Likewise, in the case of the 12-month SPI, MLR performed the best in 28 out of the 48 scenarios, while ANN and XGBR yielded the best estimations for 14 and 6 scenarios, respectively. Lastly, for the 24-month SPI, MLR achieved the best performance in 27 scenarios, while ANN and the XGBR obtained best estimations in 15 and 6 scenarios, respectively. Thus, the XGBR model showed promising results and performed as well as ANN and MLR for the test dataset, while outperforming them for the train dataset. The KNN model, on the other hand, exhibited robust performances for the train dataset, whereas it performed weaker for the test dataset compared to other ML models, making it the least favourable choice. Future studies can explore alternative ML models to further improve the accuracy of SPI predictions. Furthermore, the study found that the optimal number of input variables (lead times) varied for each SPI and station. Further studies are required to investigate further into the right selection of lead times through data statistical analyses of SPI values.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00704-023-04710-y>.

Author contribution Conceptualization, G.E.T., T.H. and M.N.; methodology, M.N., R.P., G.E.T., T.H. and F.G.; software, M.N. and R.P.; validation, R.P. and M.N.; formal analysis, M.N.; investigation, M.N.; resources, G.E.T., M.N. and T.H.; data curation, G.E.T., T.H. and F.G.; writing—original draft preparation, M.N., R.P., G.E.T., T.H. and F.G.; writing—review and editing, M.N., R.P., G.E.T., T.H. and F.G.; visualization, R.P. and F.G.; supervision, S.H.A.; project administration, M.N.

Data availability The data can be shared upon reasonable request to the corresponding author.

Declarations

Ethical approval N/A.

Consent to participate N/A.

Consent for publication N/A.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

- Abedi R, Costache R, Shafizadeh-Moghadam H, Pham QB (2022) Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto Int* 37:5479–5496
- Abeyingha NS, Rajapaksha URLN (2020) SPI-based spatiotemporal drought over Sri Lanka. *Adv Meteorol* 2020:9753279. <https://doi.org/10.1155/2020/9753279>
- Adikari KE, Shrestha S, Ratnayake DT, Budhathoki A, Mohanasundaram S, Dailey MN (2021) Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. *Environ Model Softw* 144:105136
- Almedej J (2014) Drought analysis for Kuwait using standardized precipitation index. *Sci World J* 2014:451841. <https://doi.org/10.1155/2014/451841>
- Belayneh A, Adamowski J (2012) Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *App Comput Intell Soft Comput* 2012:6–6
- Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J Hydrol* 508:418–429
- 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
- Bisong E (2019) Building machine learning and deep learning models on Google cloud platform. *Apress, Berkeley*, pp 59–64
- Caloiero T (2018) SPI trend analysis of New Zealand applying the ITA technique. *Geosciences* 8(3):101
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp 785–794
- Choubin B, Khalighi-Sigaroodi S, Malekian A, Kişi Ö (2016) Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrol Sci J* 61:1001–1009
- Dabanlı İ, Şen Z, Yeleğen MÖ, Şişman E, Selek B, Güçlü YS (2016) Trend assessment by the innovative-Şen method. *Water Resour Manage* 30:5193–5203
- Dikshit A, Pradhan B, Santosh M (2022) Artificial neural networks in drought prediction in the 21st century—a scientometric analysis. *Appl Soft Comput* 114:108080
- El Ibrahimy A, Baali A (2018) Application of several artificial intelligence models for forecasting meteorological drought using the standardized precipitation index in the Saïss Plain (Northern Morocco). *Int J Intell Eng Syst* 11:1
- Elbeltagi A, Kumar M, Kushwaha NL, Pande CB, Dittthakot P, Vishwakarma DK, Subeesh A (2023a) Drought indicator analysis and forecasting using data driven models: case study in Jaisalmer, India. *Stoch Env Res Risk Assess* 37:113–131
- Elbeltagi A, Pande CB, Kumar M, Tolche AD, Singh SK, Kumar A, Vishwakarma DK (2023b) Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-023-25221-3>
- Fadaei-Kermani E, Barani GA, Ghaeini-Hessaroeyeh M (2017) Drought monitoring and prediction using K-nearest neighbor algorithm. *J AI Data Min* 5:319–325
- Farrokhi A, Farzin S, Mousavi SF (2020) A new framework for evaluation of rainfall temporal variability through principal component analysis, hybrid adaptive neuro-fuzzy inference system, and innovative trend analysis methodology. *Water Resour Manage* 34:3363–3385
- GDM (General Directorate of Meteorology) (2020) Seasonal normals of the provinces. Retrieved from <https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx>. Accessed 1 Jan 2023
- Ghazipour F, Mahjouri N (2022) A multi-model data fusion methodology for seasonal drought forecasting under uncertainty: application of Bayesian maximum entropy. *J Environ Manage* 304:114245
- Guttman NB (1999) Accepting the standardized precipitation index: a calculation algorithm. *J Am Water Resour Assoc* 35(2):311–322
- Ham Y-S, Sonu K-B, Paek U-S, Om K-C, Jong S-I, Jo K-R (2023) Comparison of LSTM network, neural network and support vector regression coupled with wavelet decomposition for drought forecasting in the western area of the DPRK. *Nat Hazards* 116:2619–2643
- Hırca T, Eryılmaz Türkan G, Niazkar M (2022) Applications of innovative polygonal trend analyses to precipitation series of Eastern Black Sea Basin. *Turkey Theor App Climatol* 147(1–2):651–667
- Hırca T, Eryılmaz Türkan G (2022) Comparison of statistical methods to graphical method in precipitation trend analysis, a case study: Coruh Basin, Turkey. *Iran J Sci Technol - Trans Civ Eng* 46(6):4605–4617
- Hosseini-Moghari SM, Araghinejad S (2015) Monthly and seasonal drought forecasting using statistical neural networks. *Environ Earth Sci* 74:397–412
- Karstarlı Ç, Kömürçü Mİ, Akpınar A, Uzlu E, Kankal M, Önsoy H (2011) Doğu Karadeniz Havzasındaki hidroelektrik potansiyelin analizi. https://www.imo.org.tr/resimler/ekutuphane/pdf/16935_09_06.pdf
- Khan MdMH, Muhammad NS, El-Shafie A (2020) Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *J Hydrol* 590:125380
- Kumar P, Reddy MM, Sushanth T, Kumar RS, Kumar TS (2023) Real time rainfall prediction for Indian states using XGBoost and Random Forest Approach. *Int J Creat Res Thoughts (IJCRT)* 11(4) e672–e679. ISSN:2320–2882. <http://www.ijcrt.org/papers/IJCRT2304571.pdf>. Accessed 1 Jan 2023
- Taesam Lee (2023) Standardized precipitation index (<https://www.mathworks.com/matlabcentral/fileexchange/26018-standardiz>)

- ed-precipitation-index), MATLAB Central File Exchange. Accessed June 6, 2023
- Maca P, Pech P (2016) Forecasting SPEI and SPI drought indices using the integrated artificial neural networks. *Comput Intell Neurosci* 2016:3868519. <https://doi.org/10.1155/2016/3868519>
- Malik A, Kumar A, Rai P, Kuriqi A (2021) Prediction of multi-scalar standardized precipitation index by using artificial intelligence and regression models. *Climate* 9:28
- Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. *Ecol Model* 198:127–138
- Mondol MAH, Ara I, Das SC (2017) Meteorological drought index mapping in Bangladesh using standardized precipitation index during 1981–2010. *Adv Meteorol* 2017:4642060. <https://doi.org/10.1155/2017/4642060>
- Niazkar M (2020) Assessment of artificial intelligence models for calculating optimum properties of lined channels. *J Hydroinfr* 22(5):1410–1423
- Niazkar M, Zakwan M (2023) Developing ensemble models for estimating sediment loads for different times scales. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-023-03263-4>
- Odemis B, Evrendilek F (2007) Monitoring water quality and quantity of national watersheds in Turkey. *Environ Monit Assess* 133:215–229
- Pande CB, Costache R, Sammen SSh, Noor R, Elbeltagi A (2023) Combination of data-driven models and best subset regression for predicting the standardized precipitation index (SPI) at the Upper Godavari Basin in India. *Theoret Appl Climatol* 152:535–558
- Piraei R, Afzali SH, Niazkar M (2023) Assessment of XGBoost to Estimate Total Sediment Loads in Rivers. *Water Resources Management*. <https://doi.org/10.1007/s11269-023-03606-w>
- Prodhan FA, Zhang J, Sharma TPP, Nanzad L, Zhang D, Seka AM, Ahmed N, Hasan SS, Hoque MZ, Mohana HP (2022) Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach. *Sci Total Environ* 807:151029
- Şen Z (2012) Innovative trend analysis methodology. *J Hydrol Eng* 17(9):1042–1046
- Spennemann PC, Rivera JA, Saulo AC, Penalba OC (2015) A comparison of GLDAS soil moisture anomalies against standardized precipitation index and multisatellite estimations over South America. *J Hydrometeorol* 16:158–171
- Taylan ED, Terzi Ö, Baykal T (2021) Hybrid wavelet–artificial intelligence models in meteorological drought estimation. *J Earth Syst Sci* 130:38
- Vodounon RB, Soude H (2022) Drought forecasting in Alibori Department in Benin using the standardized precipitation index and machine learning approaches. *Int J Adv Comput Sci Appl* 13(12). <https://doi.org/10.14569/IJACSA.2022.01312113>
- Yılmaz B (2019) Analysis of hydrological drought trends in the GAP region (southeastern Turkey) by Mann-Kendall test and Innovative Sen Method. *Appl Ecol Environ Res* 17(2):3325–3342
- Yüksek Ö, Kankal M, Üçüncü O (2013) Assessment of big floods in the Eastern Black Sea Basin of Turkey. *Environ Monit Assess* 185(1):797–814. <https://doi.org/10.1007/s10661-012-2592-2>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.