IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 26, 2021, accepted April 7, 2021, date of publication April 20, 2021, date of current version May 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074305

Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms

ALI MOKHTAR[1,2], MOHAMMADNABI JALALI[3], HONGMING HE[1,4], NADHIR AL-ANSARI[5], AHMED ELBELTAGI[6,7], KARAM ALSAFADI[8], HAZEM GHASSAN ABDO[9], SAAD SH. SAMMEN[10], YEBOAH GYASI-AGYEI[11], AND JESÚS RODRIGO-COMINO[12,13]

[1]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest Agriculture and Forestry University, Chinese Academy of Sciences and Ministry of Water Resources, Yangling 712100, China
[2]Department of Agricultural Engineering, Faculty of Agriculture, Cairo University, Giza 12613, Egypt
[3]Department of Water Engineering, Aburaihan Campus, University of Tehran, Tehran 1417466191, Iran
[4]School of Geographic Sciences, East China Normal University, Shanghai 210062, China
[5]Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, 97187 Luleå, Sweden
[6]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China
[7]Department of Agricultural Engineering, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt
[8]School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
[9]Geography Department, Faculty of Arts and Humanities, Tartous University, Tartous 51003, Syria
[10]Department of Civil Engineering, College of Engineering, University of Diyala, Baqubah 00964, Iraq
[11]School of Engineering and Built Environment, Griffith University, Nathan, QLD 4111, Australia
[12]Department of Physical Geography, University of Trier, 54296 Trier, Germany
[13]Soil Erosion and Degradation Research Group, Department of Geography, Valencia University, 46010 Valencia, Spain

Corresponding authors: Hongming He (hongming.he@yahoo.com) and Nadhir Al-Ansari (nadhir.alansari@ltu.se)

This work was supported in part by the Second Tibetan Plateau Scientific Expedition and Research Program under Grant SQ2019QZKK2003, in part by the National Key Research and Development Program under Grant 2017YFC0505200 and Grant 2017YFC0505205, in part by the Project of the Integrated Scientific Expedition of the Ailao-Wuliang Mountains National Park under Grant 2019IB018, in part by the National Natural Science Foundation of China under Grant 41672180, in part by the Key Platforms and Scientific Research Projects in Universities in Guangdong Province of China under Grant 2018KTSCX212, and in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA23020603 and Grant XDA230000000.

ABSTRACT Accurate estimation of drought events is vital for the mitigation of their adverse consequences on water resources, agriculture and ecosystems. Machine learning algorithms are promising methods for drought prediction as they require less time, minimal inputs, and are relatively less complex than dynamic or physical models. In this study, a combination of machine learning with the Standardized Precipitation Evapotranspiration Index (SPEI) is proposed for analysis of drought within a representative case study in the Tibetan Plateau, China, for the period of 1980-2019. Two timescales of 3 months (SPEI-3) and 6 months (SPEI-6) aggregation were considered. Four machine learning models of Random Forest (RF), the Extreme Gradient Boost (XGB), the Convolutional neural network (CNN) and the Long-term short memory (LSTM) were developed for the estimation of the SPEIs. Seven scenarios of various combinations of climate variables as input were adopted to build the models. The best models were XGB with scenario 5 (precipitation, average temperature, minimum temperature, maximum temperature, wind speed and relative humidity) and RF with scenario 6 (precipitation, average temperature, minimum temperature, maximum temperature, wind speed, relative humidity and sunshine) for estimating SPEI-3. LSTM with scenario 4 (precipitation, average temperature, minimum temperature, maximum temperature, wind speed) was relatively better for SPEI-6 estimation. The best model for SPEI-6 was XGB with scenario 5 and RF with scenario 7 (all input climate variables, i.e., scenario 6 + solar radiation). Based on the NSE index, the performances of XGB and RF models are classified as good fits for scenarios 4 to 7 for both timescales. The developed models produced satisfactory results and they could be used as a rapid tool for decision making by water-managers.

INDEX TERMS Drought events, SPEI, machine learning, Extreme Gradient Boost, Tibetan Plateau.

I. INTRODUCTION

Drought is an environmental disaster that can devastate regional agriculture, water resources and ecosystem services as well as human settlements [1]–[3]. Globally, the frequency and intensity of extreme drought events are expected to increase [4]. Human impacts on the atmospheric dynamics could be considered as one of the main reasons for the increasing severity, frequency and extent of droughts during the recent decades [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

Extreme droughts adversely impact on water resource imbalance through excessive evapotranspiration and moisture deficiency [6]. Some researchers have demonstrated that drought can result in unaffordable socioeconomic losses, agriculture productivity reduction, and ecosystem degradation [7]–[10]. China has experienced severe drought periods during the last few decades that have resulted in agricultural production reduction and large losses in the economic sectors [11], [12]. For instance, in 2011 drought episodes affected 16.3 million people, and were assessed as having caused a direct economic loss of about 15.8 billion dollars (*http://www.mwr.gov.cn/sj/tjgb/zgshzhgb/201612/t20161222_776089.html*). In order to mitigate the negative effects of extreme droughts, it is imperative to develop an efficient early drought warning tool that can be valuable for mitigating rural community devastation. Based on previous studies, the Penman-Monteith (PM) equation was recommended as the sole standard method by the Food and Agriculture Organization (FAO) to calculate evapotranspiration. Further, it is one of the main components used to estimate the Standardized Precipitation Evaporation Index (SPEI) drought index based on the water balance equation. However, a significant limitation of using FAO56-PM is in its high number of meteorological variables as inputs and high uncertainties. The dynamic nature of meteorological variables is non-linear, nonstationary, and highly stochastic, making it difficult for models to be accurate. Therefore, evapotranspiration as a complex phenomenon requires accurate methods beyond empirical models.

It is very important to find sustainable solutions to manage the water scarcity under global warming and accurately predict drought occurrences [13], [14]. Drought forecasts can be done using either physical/conceptual or data driven models. While physical/conceptual models are good at providing insights into the catchment processes, they have been criticized for being difficult to adopt for forecasting applications, requiring many different types of data and resulting in models that are overly complex [15] .

In contrast, data driven models have minimum information requirements and rapid development time, being also accurate for various hydrological forecasting applications [16] . With the development of computer technology, machine learning (ML), which is a subset of artificial intelligence (AI), models have been applied in hydrology to deal with complex phenomena [17]–[19], including drought prediction [20]–[23]. Furthermore, the use of ML tools shows outstanding performance, being less time consuming and providing good accuracy [10] , [24]–[28]. The flexibility and adaptability of AI could make it useful for predicting the occurrence of droughts that occur with varying durations, frequencies and intensities. These characteristics of droughts are not effectively using empirical relations.

To date, diverse ML models such as support vector machine (SVM) [29], [30], artificial neural networks (ANN) [31], [32] and Extreme Learning Machine (ELM) [33], have been employed. Among them, SVM is the most widely used for drought prediction [34]–[36]. For example, a SVM model was combined with SPEI to predict drought over Pakistan [5] and with the Palmer Drought Severity Index (PDSI) to predict drought over Turkey [37]. Other authors have also reported the higher accuracy of SVM over the ANN algorithm in predicting Standardized Precipitation Index (SPI) over Iran [38]. Among 5 ML models, SVM emerged as the best model for estimating Combined Terrestrial Evapotranspiration Index (CTEI) in the Ganga (Ganges) River basin in India [39]. ANN was applied to analyse monthly SPEI in Kansas, USA, during 1980-2010. The confusion matrices from the analyses suggested that ANN better predicted class membership from the SPEI than the discriminant analysis for prediction of grass prairies production [40]. Further, ANN provided a greater accuracy than multiple linear regression (MLR) in forecasting the 1, 3, 6 and 12 months SPEI in Wilsons Promontory in Victoria, Australia [41]. By contrast, effective Drought Index (EDI) in eastern Australia was better predicted by extreme learning machine (ELM) over the ANN model for the overall test sites [42]. The investigation of [14] applied the Wavelet-ARIMA-ANN (WAANN) and the Wavelet-Adaptive Neuro-Fuzzy Inference System (WANFIS) models to predict the SPEI at the Langat River Basin in Malaysia for 1-month, 3- months and 6 months aggregation timescales. The results concluded that the WAANN model could be better for the prediction of SPEI-3 and SPEI-6.

Most of the artificial intelligence models are highly accurate but they are complex and attract high computational costs through the training phase. On the contrary, Rule-based Decision Tree (DT) and tree-based ensemble methods, e.g., Gradient Boosting (GB), Extreme Gradient Boosting (XGB) and Random Forest (RF), are increasingly becoming attractive because they are considered simpler, powerful and robust predictive algorithms [43]–[45]. Recently, the XGB was applied for predicting the meteorological drought in Shaanxi province, China, with promising results [36]. Deep learning (DL) models, a subset of ML, are now widely used with high accuracy. Therefore, deep learning is gaining much attention in several research fields [46]–[49]. For time series predictions, DL models of the long short-term memory (LSTM) [50], [51] and the one-dimensional convolutional neural network (1D CNN) [52], [53] are found to be more appropriate. However, drought monitoring using DL is relatively scarce [54]. For example, one study investigated SPEI-1 and SPEI-3 drought index prediction based on the LSTM model over the New South Wales region of Australia during 1901-2018 [55]. This study used precipitation, vapor pressure, cloud cover, potential evapotranspiration and temperature to predict SPEI. In the meantime, LSTM architecture was developed to forecast SPEI in Australia based on hydrometeorological and climatic variables as predictors [56]. Moreover, LSTM was applied to forecast SPEI during the period of 1958-2014 [57]. Therefore, the use of LSTM in drought forecasting is still in its infancy. Deep feed forward neural network modeling was applied to predict SPEI

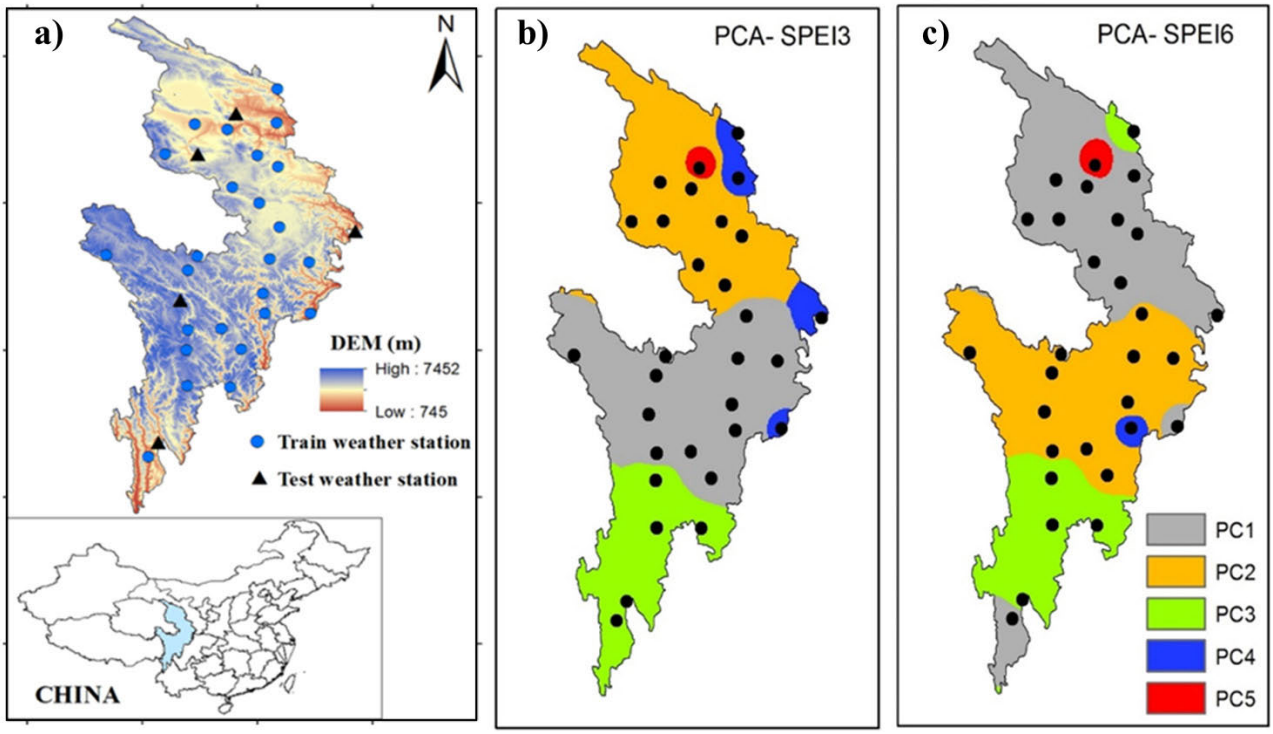

FIGURE 1. The study area, data management scenarios and training/test split used (a), principal component analysis (PCA) for SPEI-3 (b) SPEI-6 (c).

based on multi-source remote sensing data in Wuhan, China, during the period of 2001-2013 [54].

In this proposed work, the issue addressed is based on the need to estimate drought in order to minimize its damaging effects at the regional and local levels. In this regard, two DL models (CNN and LSTM) were used as rapid decision tools for SPEI modeling for two different timescales at the regional and local levels. To the best of our knowledge, the applied approaches are still poorly investigated for drought studies, especially those based on different climate inputs at both regional and local levels. Therefore, the objectives of this research are to 1) identify drought events and describe their temporal and spatial characteristics (total drought duration, severity and intensity) based on 30 meteorological stations during 1980-2019, 2) develop an alternative model (RF, XGB, CNN and LSTM) to simulate SPEIs and 3) compare the accuracy and stability of these models in SPEI predictions and to select the best model based on the prediction accuracy. The novel contribution of this work is to develop and validate the usefulness of the CNN and LSTM architecture for SPEI estimating at different timescales in China. Thus, this investigation presents a pioneering modeling strategy that would lead to improvement of efforts to address the deficiencies in drought prediction, which in turn may assist in mitigation plans such as policies for sustainable water use and the management of water supply systems.

II. STUDY AREA

The study area (419,000 km^2) is located in the eastern edge of the Qinghai-Tibet Plateau which is connected to the Loess Plateau, Sichuan Basin, and Yunnan-Guizhou Plateau (Figure 1). The Qinghai–Tibet Plateau is one of the most sensitive areas to global climate change [58], [59], and has an average altitude of about 4000 m [60]. Since the 1990s, there has been accelerated temperature rise on the Tibetan Plateau, resulting in the rapid melting of glaciers and permanent snow cover, rising snowlines, reduced wetland area, and a deteriorating ecological environment [61]. The study area has a typical alpine climate regime with an annual average temperature of 4.8°C and an annual average precipitation of 693.2 mm. More than 80% of the annual precipitation occurs during the wet period from May to August. The variability of drought in China is influenced by the anomaly of atmospheric circulation such as the East Asian summer monsoon (EASM) [62] and EI Niño-Southern Oscillation (ENSO) [63]. Soil freezing and snow cover usually occur from early October to the end of April [64]. The region is rich in snow and glacier resources and is the source of many rivers in Asia, including the Yellow, Yangtze and Lancang Rivers. Due to global warming, accelerated melting of glaciers and the redistribution of precipitation lead to frequent floods and snowstorms over the Tibetan Plateau [65].

III. MATERIAL AND METHODS

Figure 2a shows the workflow of this research as explained in the ensuing.

A. DATA SOURCES

Climate datasets at daily timescale from 30 meteorological stations for the period of 1980–2019 were retrieved from

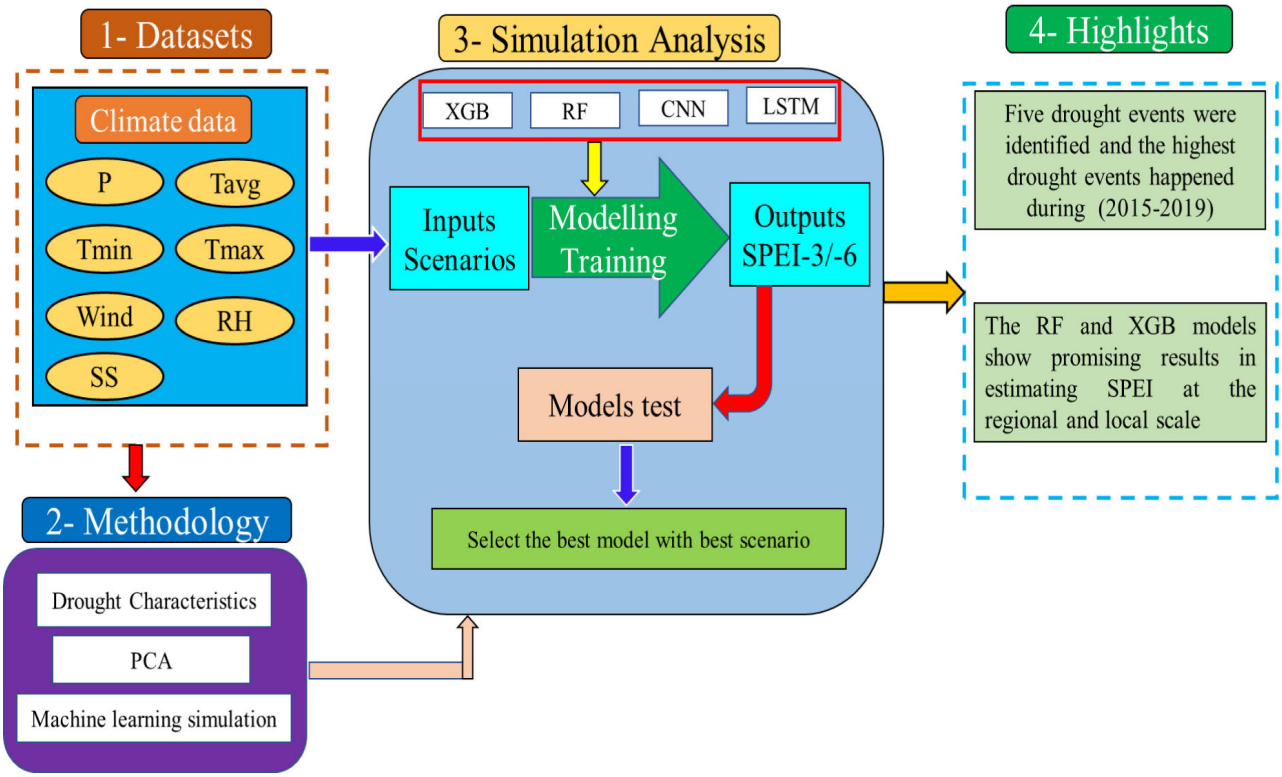

FIGURE 2. Flowchart of the research methodology.

the China National Meteorological Data Sharing Platform (*http://data.cma.cn/en*). The climate variables were daily precipitation amount (mm), temperature (minimum, maximum, and average (°C)), solar radiation (Wm^{-2}), sunshine hours (h), wind speed at 2 m (ms^{-1}), and relative humidity (%). Based on previous studies, the PM equation was recommended as the sole standard method by FAO to calculate reference evapotranspiration (ET_0) and it has been successfully applied in China. It incorporates both physiological and meteorological parameters and has been widely used globally as a result of its intrinsic rationality and reliability [66]–[69]. The detailed steps for the calculation of ET_0 were reported in the investigation of Mokhtar *et al.* (2020) [66]. Therefore, the PM equation based on the climate variables was used to calculate the monthly ET_0 over the study area. To develop SPEI indices at different timescales, the monthly climatic water balance series was aggregated with an n-month moving sum, i.e., the current monthly value plus the previous n-1 monthly values; n = 3 corresponds to SPEI-3 and n = 6 for SPEI-6. This means that the SPEI-3 value for March is the sum of monthly SPEI values for January–February–March [70].

B. SPEI CALCULATION

With over 150 drought indices in the literature, validating each one and developing a common consensus is not feasible. However, there seems to be a growing consensus about the use of SPEI in recent times, primarily because it uses both rainfall and temperature to determine it rather than only rainfall as in the case of the Standard Precipitation Index (SPI) [55], [71]. Moreover, the SPEI is one of the widely used drought indices for tracking drought evolution over different timescales of interest (i.e., 1, 3, 6, 9, 12, and 24 months) [70]. The SPEI uses the difference between precipitation and evapotranspiration to represent the regional drought [72], [73]. Thus, SPEI was selected as the drought index considering the effect of temperature-induced evapotranspiration on droughts.

To calculate the SPEI, the monthly water balance (D_i) which is the difference between the monthly precipitation (P_i) and monthly potential evapotranspiration (PET_i) is first calculated. These values are then aggregated at the timescales of interest [74] as:

$$D_i = P_i - PET_i \tag{1}$$

$$D_n^k = \sum_{i=0}^{k-1} (P_{n-i} - PET_{n-i}), \quad n \geq k \tag{2}$$

where k = 6 for SPEI-6 is the aggregation timescale and n is the 6^{th} month for June. A three-parameter log-logistic probability distribution was used to fit the D series. The cumulative distribution function F(x) is given as:

$$F(x) = \left[1 + \left(\frac{\alpha}{x - \gamma}\right)^{\beta}\right]^{-1} \tag{3}$$

where α, β and γ are the scale, shape, and location parameters, respectively.

The SPEI value is obtained as the standardized value given as:

$$SPEI = W - \frac{c_0 + c_1 W + c_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \quad (4)$$

$$W = \sqrt{-2\ln(P)} \quad for \quad P \leq 0.5, \quad (5)$$

where P is the probability of exceedance of a given D value, the cumulative distribution of D being F(D). When $P > 0.5$ it is replaced with the non-exceedance probability (F(D) $= 1 - P$) and the sign of the calculated SPEI is reversed. The constants are $c_0 = 2.5155$, $c_1 = 0.8028$, $c_2 = 0.0103$, $d_1 = 1.4328$, $d_2 = 0.1892$ and $d_3 = 0.0013$. Generally, SPEI can be calculated at different time scales of 1, 3, 6, 9, or 12 months [75]. However, we restricted our study to only SPEI-3 and SPEI-6 representing 3 and 6 monthly timescales, respectively, which are considered as representative of agricultural drought where precipitation shortage in a short period is the main cause of adverse drought effects [76]. The SPEI category classification is presented in Table 1 based on the Agnew's scheme [77]. In this study, we are following a different SPI/SPEI classification presented in the literature (e.g., [78], [79]. However, we used a different classification for drought intensity as proposed by Agnew *et al.* [77]. It is based on the SPI as a meteorological drought that is similar to SPEI. Agnew's approach uses probability classes rather than the quantile values of the index for classification. Here, the impacts are most noticeable as it allows identification of mild and moderate droughts.

TABLE 1. The Agnew's scheme for drought categories classification (Agnew, 2000).

SPEI values	Drought category
> 0	No drought
0 to -0.5	No drought
-0.5 to -0.84	Moderate drought
-0.84 to -1.28	Severe drought
-1.28 to -1.65	Extreme drought
$<$ -1.65	Very extreme drought

C. DROUGHT CHARACTERISTICS

Several methods have been widely used for drought event identification, including the threshold level method, run theory, and empirical orthogonal functions [80]. For example, a drought event is defined as a period in which the SPI is continuously negative and can reach a value of -1.0 or less according to McKee *et al.* [81], [82]. In this study we determined the drought events based on a threshold value of SPEI < 0 (i.e. all negative values considered), and also on a threshold value of SPEI < -0.84 which is more suitably for drought category. Agnew's approach uses probability classes rather than the quantile values of the index for classification. Here, the impacts are most noticeable as it allows identification of mild and moderate droughts. The drought characteristics of interest for this study are the Total Drought Duration (TDD), Drought Severity (DS), drought peak (DP) and Spatial Extent of Drought (SEoD), which were analyzed for specific drought events. A drought event was defined as being when the SPEI values were negative and also when the SPEI were less than -0.84 [81]. Therefore, the total duration of a drought episode (TDD) is defined as either the sum of the number of continuous months n_i when the SPEI $<$ 0 [83], [84], [82], or when the SPEI is less than -0.84 (less than moderate drought), as such, the TDD was expressed for all drought episodes for a case where SPEI is <0 (TDD_0) and also for a case where SPEI is less than -0.84 ($TDD\text{-}_{0.84}$) over the whole studied period N_i as:

$$TDD(\%) = \sum \frac{n_i}{N_i} \times 100 \quad (6)$$

where n_i is the sum ofmonthsencompassingdrought episodes, N_i isthetotalstudied period in months for each station, and TDD being either TDD_0 or $TDD\text{-}_{0.84}$. The DS is the absolute sum of SPEI values during a specific duration DD and calculated when SPEI is less than -0.84 as:

$$DS = \left| \sum_{i=1}^{DD} SPEI_i \right| \quad (7)$$

$SPEI_i$ represents the SPEI value in month i. Moreover, DP refers to the minimum SPEI value during the drought event [85]. The drought-prone areas were examined by the percentage of the number of drought locations in the total study area for the different drought categories. The percentage of area affected by drought (SEoD) was determined [79], [81], [87] as:

$$SEoD(\%) = \frac{m}{M} \times 100 \quad (8)$$

where m is the number of stations when SPEI < 0, or SPEI < -0.84, and M is the total number of stations [81], [86].

D. PRINCIPAL COMPONENT ANALYSIS (PCA)

In order to identify the sub-regional patterns of drought covariability in the Qinghai-Tibet Plateau, we performed a principal component analysis (PCA) in an S- mode for both SPEI-3 and SPEI-6, which often has been used to obtain homogenous spatial-temporal patterns of climate including droughts (e.g., [80] , [87], [88]). This multi-variate statistic can regionalize spatial patterns of drought, which reflects the high temporal variability of climate in the Qinghai-Tibet Plateau, especially at the local and the regional scales. As such, the PCA was employed to produce a few new representative SPEI data from the original one and these are

TABLE 2. Combination of input climate variables for the 7 scenarios used for the machine and deep learning models.

Scenario	P	Tave	Tmax	Tmin	Wind	RH	SH	SR
1	✓	✓						
2	✓	✓	✓					
3	✓	✓	✓	✓				
4	✓	✓	✓	✓	✓			
5	✓	✓	✓	✓	✓	✓		
6	✓	✓	✓	✓	✓	✓	✓	
7	✓	✓	✓	✓	✓	✓	✓	✓

called the principal components (PCs) [84]. PCA can result in a few linearly unhomogenized PCs to explain most of the temporal variance from the main SPEI data. The PCA is used in climatology as Empirical Orthogonal Function (EOF) analysis [85]. Such a transformation is a linear one and depends on the eigenvectors of a correlation or covariance matrix [86], [88] . The study employed a correlation matrix with the eigenvalues and eigenvectors using a weighted coefficient. After the results of the EOFs, we retained only the significant PCs which resulted in an eigenvalue higher than unity [89]. The final eigenvectors, which are called ''loadings'', represent the correlation between the time series of SPEI and the PCs scores. Ultimately, PCA reduces SPEI time series dimensions with maximum total variance of SPEI data and find localized drought patterns. The Varimax orthogonal rotation methods were applied on the retained PCs. Each SPEI time series was then set to a specific component based on the maximum correlation observed between each SPEI time series and all the PCs, 0.6 being used as the threshold standard criterion [81], [90] to split the Qinghai-Tibet Plateau into sub-regional patterns.

E. MACHINE LEARNING MODELS

Based on the results of the principal components analysis (PCA), five climate zones were identified. At the regional scale, the highest loaded station in each zone was considered to be the test station, while the remaining 25 stations constituted the model training datasets (Figure 1a). Therefore, model development and testing were based on all identified climatic zones of the region to produce a more practical model at the regional scale, and not only for a specific area or station. For the local scale, the performances of the best models in the regional scale were evaluated separately in the five PCA zones (local scale). Seven scenarios, consisting of various combinations of the climatic variables (Table 2), were investigated for their effects on SPEI-3 and SPEI-6 to provide an accurate estimation for the area with data scarcity. This may help improve drought analysis for data scarcity regions, and also to evaluate the contribution of each climate variable to the SPEI.

The SPEI timeseries was first calculated using (4) that involves precipitation and reference evapotranspiration for the establishment of the drought events and their characteristics. Then the models were used to estimate the SPEI with precipitation data and the various combination scenarios of the climatic variables (temperature, relative humidity, wind speed, sunshine and solar radiation) data as inputs (Fig. 2). To facilitate the regression task, the input data were normalized to the range from 0 to 1 as:

$$X_n = \frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \tag{9}$$

where X_n is the normalized data, X_0 is the original data, while X_{min} and X_{max} are the minimum and maximum values of the original data. Performance statistics were then used to select the best model and climate variable scenarios. Scikit-learn 0.22.1 and TensorFlow 2.1.0 libraries of the Python programming language were used to run the ML models. A virtual machine from Google Cloud Platform was utilized for the computations using a graphics processing unit. The hyperparameters of the models were optimized using a grid search method for the machine learning models and random search for the deep learning models. Below are the descriptions of the models.

1) EXTREME GRADIENT BOOSTING (XGB)

The Extreme Gradient Boosting (XGB) algorithm suggested by [91] is a novel improvement of the Gradient Boosting

Machine based on regression trees. The algorithm is based on the idea of ''boosting'', which combines all the predictions of a set of ''weak'' learners to develop a ''strong'' learner through additive training strategies. The main goal of XGB is to decrease the overfitting and underfitting problems, and also to minimize computational costs. The general function for the prediction at step t is:

$$f_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \tag{10}$$

where $f_t(x_i)$ is the learner at step t, $f_i^{(t)}$ and $f_i^{(t-1)}$ are the predictions at steps t and $t-1$, respectively, and x_i are the input variables.

To avoid overfitting problems without influence on the computational speed of the model, the XGB applies the analytic expression below to evaluate the ''goodness'' of the model from the original function as:

$$Obj^{(t)} = \sum_{k=1}^{n} l(\bar{y}_i, y_i) + \sum_{k=1}^{t} \Omega(f_i) \tag{11}$$

where l is the loss function, n is the number of observations and Ω is the regularization term which is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2 \tag{12}$$

In equation (12) ω is the vector of scores in the leaves, λ represents the regularization parameter, and γ denotes the minimum loss needed to further partition the leaf node. Detailed information and the computation procedures of the XGB algorithm can be found in [91]. We applied the XGB by using 4000 trees, 20 max depth, learning rate of 0.1 and the default values of the other hyperparameters. The following sets of hyperparameters were applied: 'n' number of trees (100, 300, 500, 1000, 2000 and 4000); max depth (1, 2, 5, 10, 15, 20, 22) and learning rate (0.05, 0.1 and 0.5).

2) RANDOM FOREST (RF)

The RF model, developed by [92], is based on an ensemble of decision trees with controlled variance. The RF model has been widely used for regression and classification problems. A random forest regression is a specific type of bootstrap ensemble. It deals with random binary trees that use a subset of the observations via bootstrapping, where a random subset of the training dataset is sampled from the raw dataset and utilized to evolve the model. The detailed data and computation procedure of the RF model can be found in [93, 94]. To get the best score, a RF was trained using 4000 trees, 25 max depth and the default values of the other hyperparameters. During the tuning phase, the following hyperparameters sets and their respective values were used: n estimators (number of trees) (100, 300, 500, 1000, 2000 and 4000); max depth (1, 2, 5, 10, 15, 20, 25).

3) LONG SHORT-TERM MEMORY (LSTM)

LSTM is a special type of recurrent neural network (RNN) [94] used to handle sequential data with advantages over traditional RNN. An LSTM network contains different memory blocks which are linked through layers. Each layer includes a set of frequently connected memory pixels and three multiplicative units, namely the input, forget and output gates ([95], [96]. The architecture of the LSTM model is shown in Figure 2b.

To identify the optimal forecasting strategy, an LSTM network and LSTM layer with 100 hidden units were used, followed by a fully connected layer of size 30. The hyperparameters of the LSTM layers were kept at their default values. In the first fully connected layer, the *tanh* activation function was used, and in the last fully connected layer (output layer), the sigmoid activation function was used. The Adam training algorithm [97] was used; learning rate was set to 0. 01, batch size was set to 128, the number of training epochs was 500 and dropout rates were assessed by 0.5 (Fig. 3a).

4) CONVOLUTIONAL NEURAL NETWORK (CNN)

The convolution layers are the main difference between CNN and conventional ANN. These layers are able to perform automatic feature extraction, capturing features of the input data which are key to figure out the relationship between the inputs and output parameters. Thus, the raw data were handled without a manual feature extraction. CNN is effective in dealing with high-dimensional data based on their shared-weights architecture and translation invariance characteristics [98]. In this study, CNN with one-dimensional (1D) convolutional filters (1D CNN) was used [93], [99]. The CNN model consists of three layers, namely the input, hidden and output layers (Figure 2c). The input of a three-dimensional array is usually fed to a convolutional layer where the dimensions are represented by height, weight and the number of channels. The hidden layer includes a convolutional layer, pooling layer, and fully connected layer. The convolutional layer is set to automatically extract the characteristics at different regions/zones of the whole raw input or the intermediate characteristic maps with learnable filters and more layers can iteratively extract complex features from the last feature [98], [100]. The pooling layer turns all the values in the pooling window to a single value. Moreover, the pooling reduces the computational cost of the learning process as well as overfitting problems [101]. The fully-connected layer, which combines all local features into global features, is used to calculate the final result [98]. The detailed information can be found in [52], [100]. In this study, a convolutional layer with the following settings was used, namely 32 filters, kernel size equal to 16, stride equal to 1, padding was set to ''same'' and a rectified linear unit (ReLU) was used as the activation function. A Max-pooling layer with pool size of 3 and dropout of 0.3 was used. In the first fully connected layer, a ReLU activation function was used, and a linear activation function was used in the last fully connected layer. The training algorithm

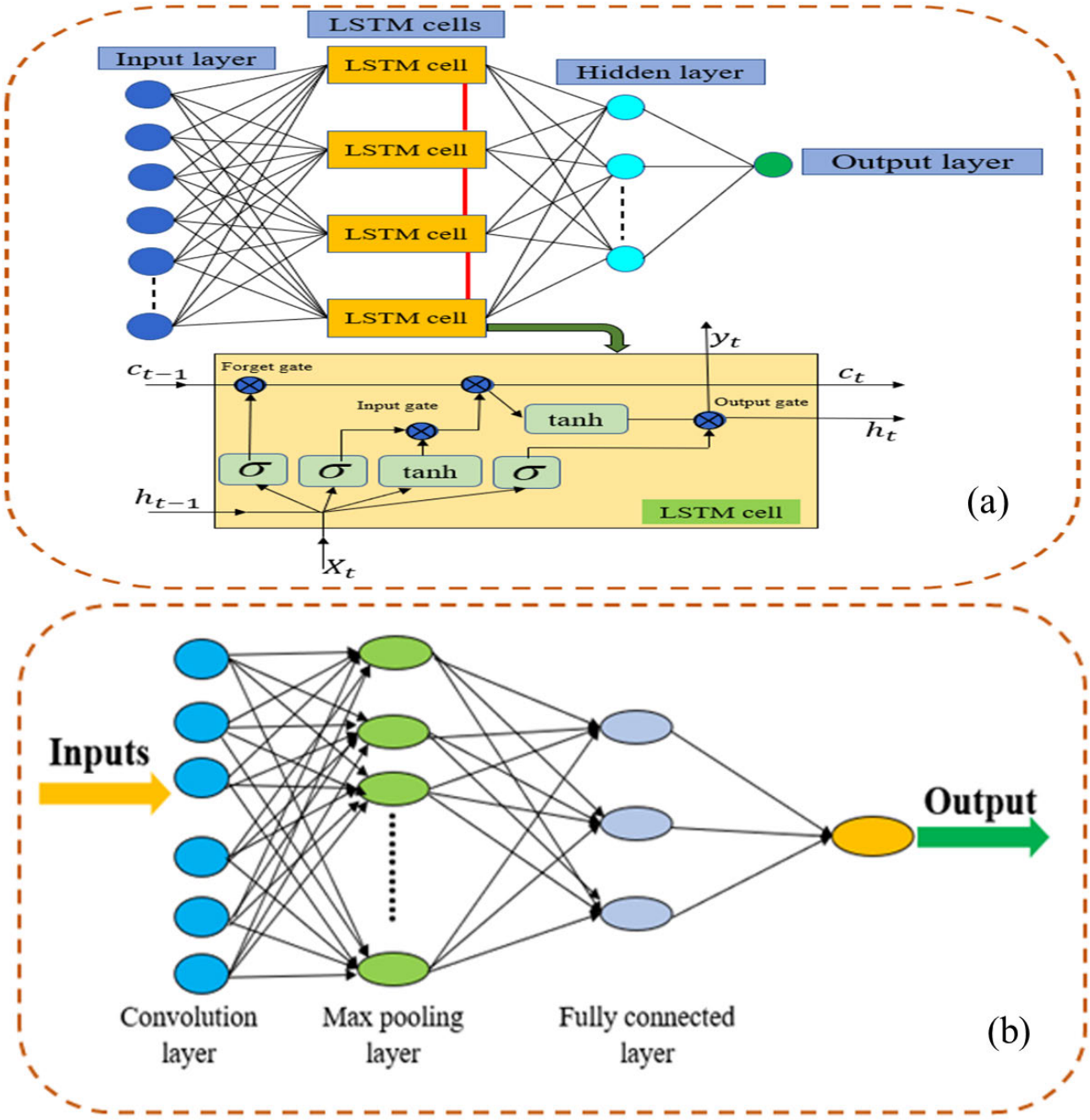

FIGURE 3. Architecture of the LSTM model and cell structure (a) and the formation of the CNN model (b).

was set with learning rate, batch size and number of training epochs of 0.001, 512 and 700, respectively (Fig. 3b).

F. PERFORMANCE STATISTICS FOR THE EVALUATION OF THE MODELS

Performance statistics of the Nash–Sutcliffe model efficiency coefficient (NSE), the mean square error (MSE), the mean absolute error (MAE), mean bias error (MBE) and correlation coefficient (R) were used to assess the applied ML models. A NSE value of 1 indicates a perfect fit, greater than 0.75 is a very good fit, between 0.64 and 0.74 is a good fit, between 0.5 and 0.64 is a satisfactory fit and less than 0.5 is an unsatisfactory fit [102]. The performance statistics are defined as:

$$NSE = 1 - \frac{\sum (P_i - O_i)^2}{\sum (\overline{O} - O_i)^2} \tag{13}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |O_i - P_i| \tag{14}$$

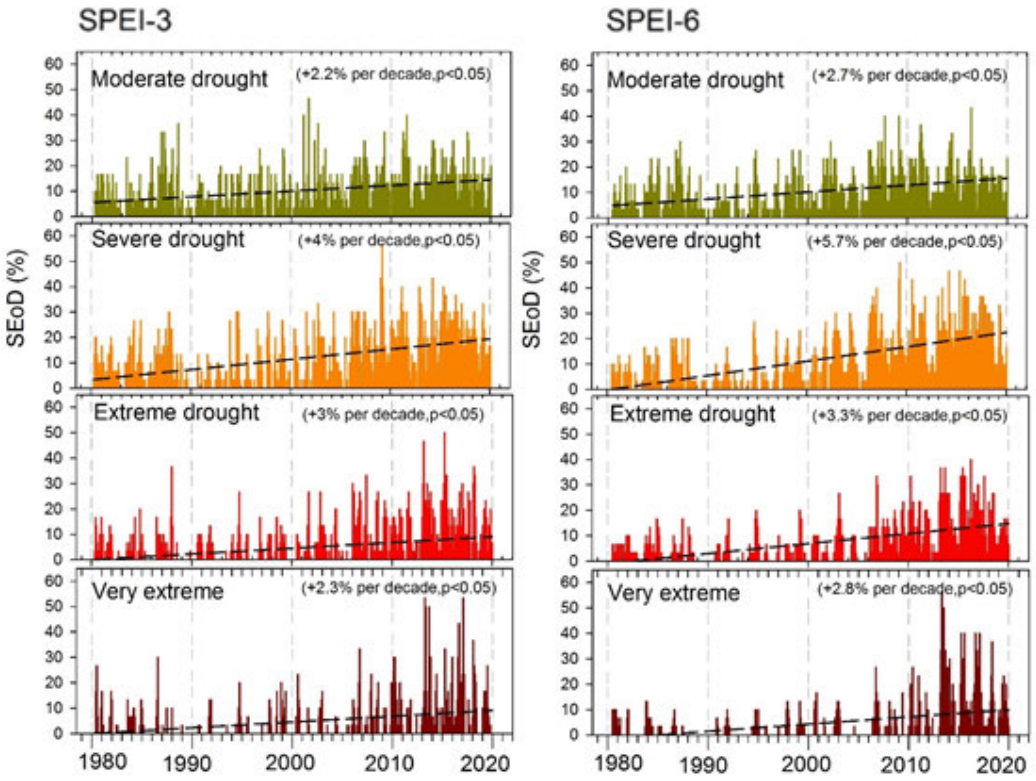

FIGURE 4. Temporal evolution of spatial extent of drought (SEoD), i.e., the percentage of stations affected by different drought categories. The amount of changes ()perdecade and trend given in dashed lines is statistically significant at the 95% level ($p < 0.05$).

$$MSE = \frac{1}{n}\sum (P_i - O_i)^2 \tag{15}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n} (O_i - P_i) \tag{16}$$

$$R = \frac{\sum_{i=1}^{n} (O_i - \overline{O})(\mathrm{P}_i - \overline{\mathrm{P}})}{\sqrt{\left(\sum_{i=1}^{n} (O_i - \overline{O}_i)^2\right)\left(\sum_{i=1}^{n} (\mathrm{P}_i - \overline{\mathrm{P}})^2\right)}} \tag{17}$$

where O_i and P_i are the actual and the predicted SPEIs, respectively, $\overline{O}$ represents the average values of the actual SPEI index, and n is the number of observations.

IV. RESULTS AND DISCUSSIONS

A. ANALYSIS OF THE DROUGHT EVENTS AND CHARACTERISTICS DURING 1980-2019

The eastern edge of the Qinghai-Tibet Plateau has been subjected to critical drought events that have extremely influenced the ecosystem productivity [74], [103]. In this regard, the current study aimed to identify drought events by investigating the spatial extent of drought (SEoD) of 30 meteorological stations at two aggregated timescales of 3 (SPEI-3) and 6 (SPEI-6) months. The exposure of the eastern edge of the Qinghai-Tibet Plateau to the different drought intensity is illustrated in Figure 4. As depicted in Figure 4, the trends of SEoD percentages of stations affected by all drought categories (i.e., moderate, severe, extreme and very extreme droughts) exhibited a significant increase ($p < 0.05$) during the studied period. By contrast, the SEoD for the mild drought exhibited a statistically non-significant increase for both SPEI3 and SPEI6. Interestingly, the percentage of stations/area exposed to severe and extreme droughts (i.e., from -0.84 to -1.28 and -1.28 to -1.65) presented increasingly strong changes from 1980 to 2019, in the order of 4 to 5.9% and 3 to 3.3% per decade ($p < 0.05$), respectively. A comparison between the SEoD for all drought categories suggests that the increases in the percentage of stations affected by severe and extreme droughts were higher than those of mild and moderate droughts, showing an equal increase in the percentage of stations affected by the drought for the mild and very extreme intensities.

As shown in the results presented by Figure 5, the number of stations impacted by severe drought (SPEI from -0.84 to -1.28) were almost 50% of all stations during 2009, from January to March for SPEI-3 and from March to May for SPEI-6. Notably, the number of stations impacted by very extreme drought (SPEI < -1.65) were almost 40-50% of all stations during the extraordinary drought event in 2013, from March to May for SPEI-3 and from March to July for SPEI-6. In 2015, more than 30-50% of the stations were subjected to extreme drought from March to August based

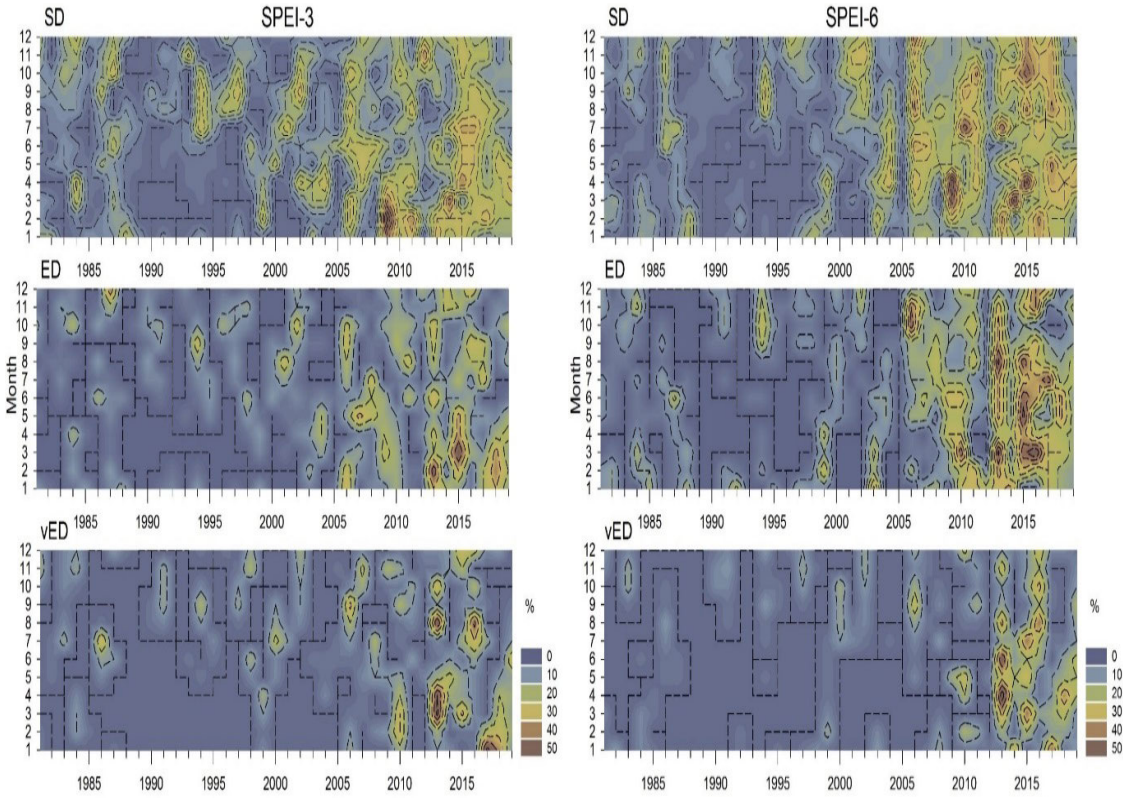

FIGURE 5. Monthly variability of total SEoD. Note: SD is severe drought, ED is extreme drought, vED and is very extreme drought.

on SPEI-6. Accordingly, the results indicated that the drought pattern tends to occur during the winter and spring seasons more than during summer and autumn. In order to accurately evaluate the spatial dimension of drought in the study area during the monitoring period of 1980-2019, four drought durations were identified randomly to detect the drought frequency and temporal evolution (D1: 2002-2004, D2: 2006-2009, D3: 2010-2014, and D4: 2015-2019) in the eastern edge of the Qinghai-Tibet Plateau and the corresponding drought characteristics (TDD, DS and DP) were analyzed based on SPEI-3 and SPEI-6. Figure 6 illustrates the spatial distribution of the drought characteristics and indicates that the D3 (2010-2014) and D4 (2015-2019) events were the most influential during the entire monitoring period of this study. In the case of D3 (2010-2014), a TDD_0 of more than 80% during that period concentrated in the southern, central and northern parts at 4 and 13 stations, respectively, for SPEI-3 and SPEI-6, 10 and 14 stations experienced severe or higher intensity drought ($TDD < -0.84$) for a long time between 40-60% of the total duration, 2 station recorded a severe or higher intensity drought by $TDD_{-0.84}$ to the tune of 60-80% for SPEI-6.

During the event D4 (2015-2019), the spatial distribution of the $TDD < 0$ events was concentrated in 80 to 100% of this period and at 12 and 18 stations for SPEI-3 and SPEI-6, respectively. With respect to $TDD < -0.84$, 6 and 12 stations experienced a severe or higher intensity drought for a long time between 60-80% of the total duration for SPEI-3 and SPEI-6, respectively. One station in the central part recorded a severe or higher intensity drought by the $TDD_{-0.84}$ measure to the value of 90-100% of the total duration for SPEI-6. Historically, drought event D4 is considered as the worst extraordinary drought event in term of intensity and duration in the Qinghai-Tibet Plateau. Also, the spatial distribution of the DS indicates concentration in the northern and central parts, ranging from 5 to 30 for both SPEIs during D1. During D3 (2010-2014) for SPEI-6, most of the study area recorded 15 to 50. By contrast, during D4 (2015-2019), the DS covered most of the study area with values more than 30% for SPEI-3 and SPEI-6. In the context of DP, the entire study area recorded TDDs < -1.65 for SPEI-3 and SPEI-6 during D3. This result was consistent with [74] who documented that severe droughts occurred in 2006 and 2011 in the Yangtze River basin. In autumn 2013, moderate and severe droughts in China occurred in most areas including Qinghai-Tibet Plateau [1], [104], [105]. In 2006 and 2011, droughts were the most severe in the Yangtze River basin, and the area affected by the droughts was the largest including the Qinghai-Tibet Plateau [74]. For example, drought areas in 2011 accounted for 54.3% of the total area of the Yangtze River basin, which was the largest drought area during the study period with 2006 experiencing the next largest drought area.

The abnormal circulation of the atmosphere caused limited water vapor transportation, which was the main cause

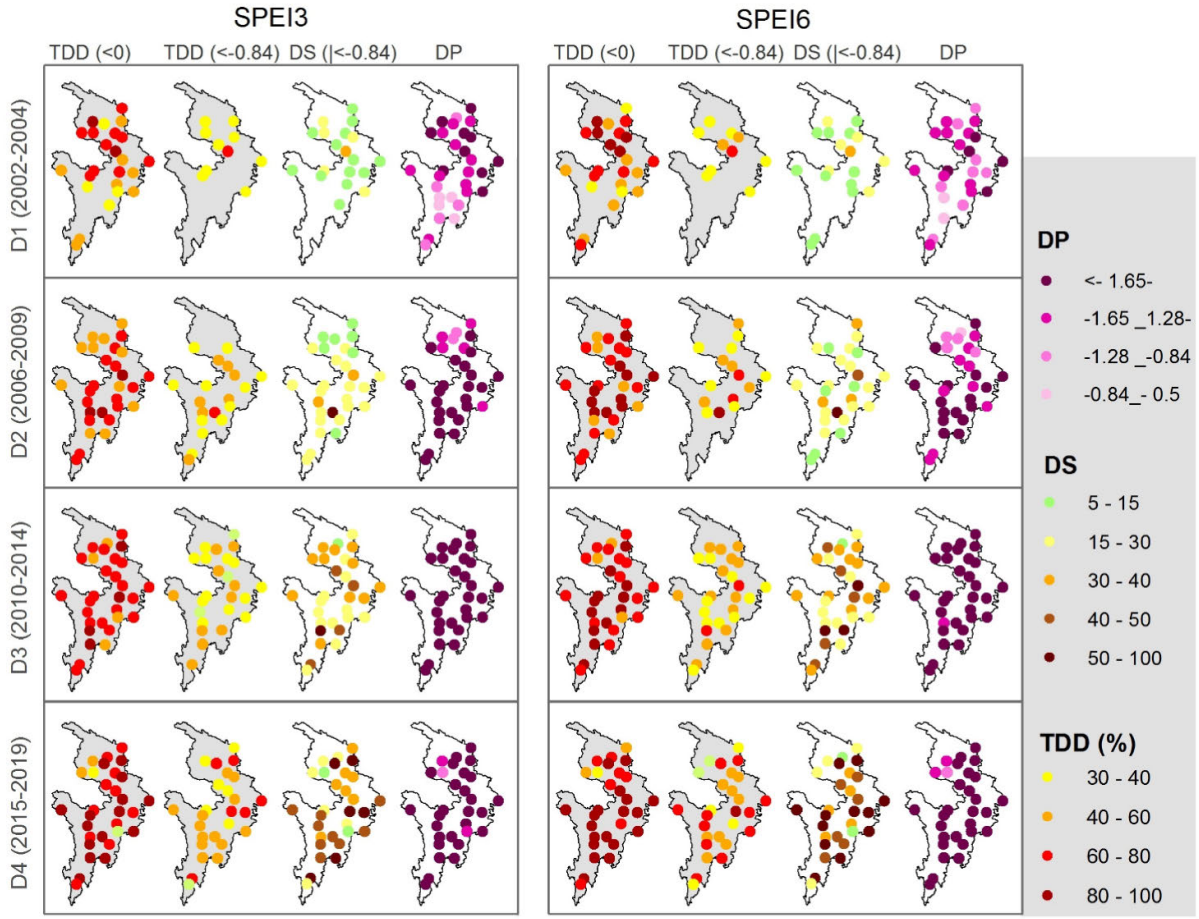

FIGURE 6. Spatial distribution of total drought duration (TDD, %), drought severity (DS), and the minimum SPEI values representing the drought peaks (DP) for the specific drought events for SPEI-3 (a) and SPEI-6 (b).

of drought and less rainfall [74]. Due to the existence of the Siberian high, the vertical movement and water vapor conditions in these areas were not conducive to precipitation [106], [107]. Thus, southwest China is one of the zones most vulnerable to climate change due to the downward airflows which are impacting on water vapor flux associated with anomalies of the atmospheric circulation [108]. Drought events during the past few decades are posing major threats to developments in many sectors over China, and have caused severe damages to many terrestrial ecosystems [109]. Nonetheless, successive drought events, namely 1997-2003 episodes [110], the 2006 episode [108], and 2009-2010 episodes [111], that hit different parts of China have caused tremendous damages especially in the agricultural sector [109]. A serious drought event in 2009 is a key example of the negative impacts of drought on the agriculture and economics. It converted part of southwest China into a desert and lakes became dried. Thus, it negatively affected aquatic animals and caused water scarcity for more than 8 million people and 3 million hectares of croplands [112], [113].

In order to more accurately detect the spatial-temporal dimension of the SPEI over the studied area, the PCA was calculated with regard to the SPEI timeseries to define the spatial patterns of drought variability in the study area. The PCs were retained for Varimax rotation in order to identify the 5 sub-regions/zones of the SPEIs (Figure 1b and c). It was observed that the first component of the PCA of the SPEIs explained the highest significant amount of variance (23-31%) but the variance seemed to be equally distributed among the 5 PCs. The first 5 components together explained about 77.2% of the total variance of SPEI-3 and 79.7% of SPEI-6, while the remaining components accounted for the rest of the total variance. But the first three PCs together explained 60.5% of the total variance for SPEI-3 and 70% for SPEI-6. Not only were the loadings corresponding to each PC mapped to show the spatial patterns of wet/dry SPEIs variability for the studied area (Figure 7), but also the temporal mode of the loadings corresponding to the extracted components was illustrated. These results suggest high heterogeneity of the evolution of the wet and dry events due to the large extent of the area that is several hundreds of kilometers across and follows several climate patterns. The strong spatial variability of drought in central China calls for drought management and strategies to be applied at the local and regional scales. Moreover, the differences in

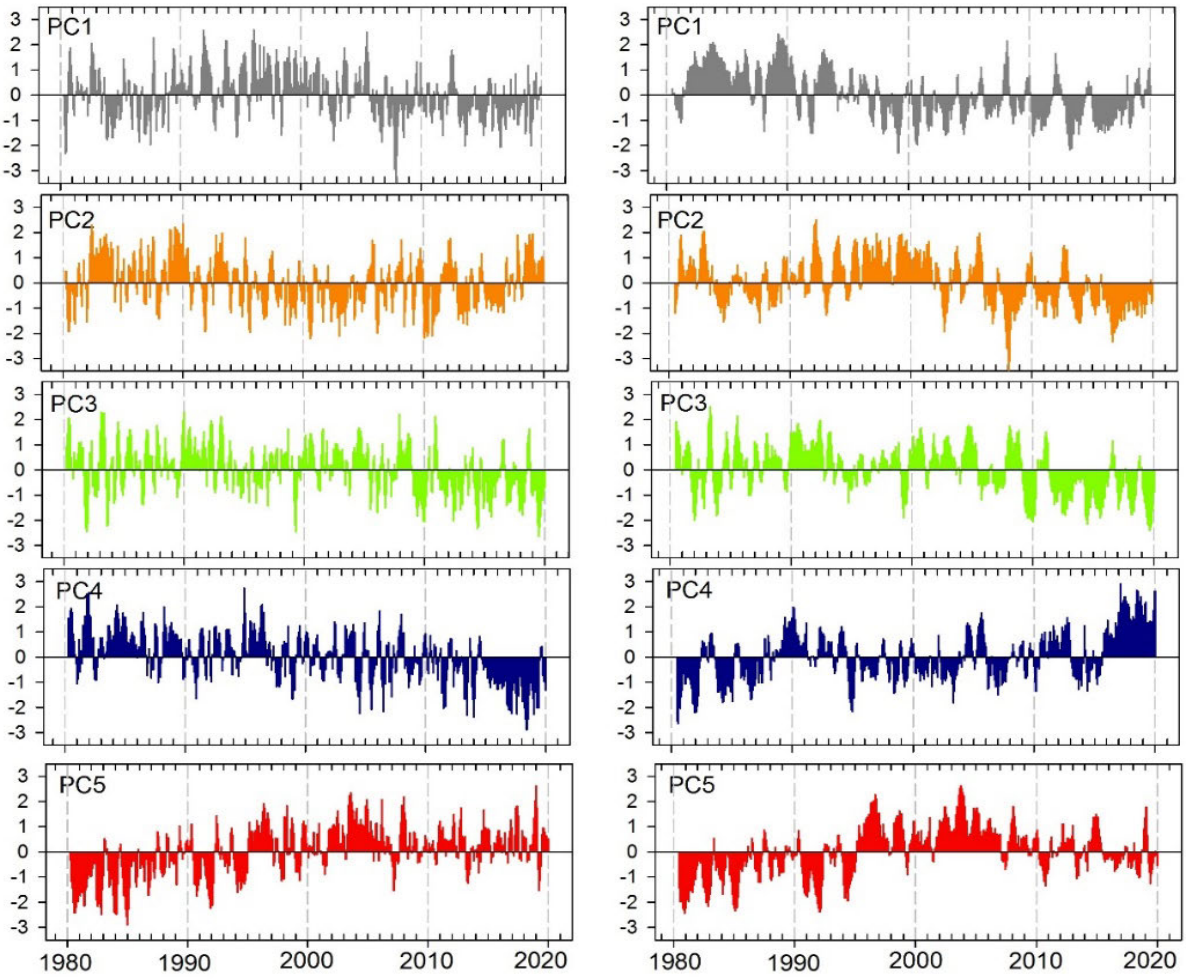

FIGURE 7. Temporal evolution and trends of the SPEI-3 and SPEI-6 from the PCs corresponding scores that represent drought patterns.

drought characteristics among the principal components are based on the changes in topography, climate, vegetation types and human factors. The combined effects of temperature and precipitation in particular is the major contributor in changes between the five zones.

B. EVALUATION OF THE MACHINE LEARNING MODELS

The performance statistics of the models using the climate data combination scenarios in predicting the SPEIs are shown in Table 3 for the two SPEIs. The MSE and MAE values for all the developed XGB-SPEI-3 models ranged from 0.26 to 0.61 and 0.39 to 0.57, respectively. The best developed XGB model for predicting drought characteristics was in combination with scenario 7, which considered all input variables (Figure 8). Considering RF for SPEI-3, scenario 6 emerged as the best (Table 3). However, under data scarcity conditions, scenario 4, which depends only on temperature and wind speed beside the precipitation data, is a good alternative for the prediction process. This scenario yielded a slight increase in MSE and the other performance metrics compared to scenario 7, but it is acceptable and satisfactory. For SPEI-6, the superiority of XGB was evident at scenario 7 that produced the lowest MSE of 0.16, MAE of 0.37, R of 0.76 and MBE of -0.09. By analyzing the XGB scenarios, it showed scenarios 4, 5, and 6 also performed well and close to scenario 7 for the XGB model. Also, RF with scenario 4 achieved good results with MSE of 0.31, MAE of 0.46 and R of 0.82. Based on the NSE index, the performance of the XGB and RF models are classified as good fits for scenarios 4 to 7 in both 3- and 6-months timescales.

The results demonstrated that scenario 4 with the RF model is good enough for assessing SPEI-6 if only temperature, wind speed and precipitation data are available. In general, XGB5 and RF6, and XGB7 and RF7 are more efficient for predicting the SPEI-3 and SPEI-6 outcomes in the Tibetan Plateau. Our results are similar to those presented by [114] who applied RF for drought forecasting for 118 stations over the period from 1963 to 2016 in southwest China. They showed MAE values of 0.36 for SPEI-3 and 0.22 for SPEI-6. Our findings also corroborate the results by [115] who assessed and monitored drought in the USA using different machine learning models. They found that the RF

TABLE 3. Performance statistics of the applied models for the seven input scenarios.

SPEI-3								
scenario	MBE				NSE			
	XGB	RF	CNN	LSTM	XGB	RF	CNN	LSTM
1	0.04	0.05	-0.07	-0.12	0.28	0.29	0.29	0.27
2	-0.01	0.03	-0.07	-0.08	0.44	0.40	0.33	0.29
3	-0.02	-0.06	-0.04	-0.03	0.69	0.61	0.40	0.39
4	-0.05	-0.08	-0.04	-0.04	0.70	0.66	0.59	0.49
5	-0.04	-0.09	-0.04	-0.04	0.70	0.57	0.59	0.48
6	-0.03	-0.08	-0.03	-0.07	0.71	0.68	0.63	0.58
7	-0.04	-0.15	-0.05	-0.08	0.71	0.67	0.67	0.53
SPEI -6								
1	0.23	-0.16	0.13	0.11	0.53	0.49	0.23	0.23
2	0.19	-0.13	0.19	0.27	0.55	0.58	0.18	0.04
3	-0.16	-0.13	0.08	0.08	0.58	0.59	0.21	0.21
4	0.08	-0.10	0.09	0.08	0.63	0.66	0.34	0.39
5	0.07	-0.11	0.04	0.01	0.64	0.63	0.36	0.40
6	0.06	-0.06	0.02	0.09	0.65	0.67	0.57	0.50
7	0.09	-0.08	0.04	0.08	0.68	0.69	0.62	0.52

model was the best with highest correlation and the lowest RMSE of 0.30, among the approaches used. This outcome is consistent with the findings presented by [115], that the accuracy of the prediction process is greater for long-term drought episodes compared to the short-term periods. Further, our finding is in agreement with the study of SPEI prediction in New South Wales (Australia) which concluded that RF achieved RMSE and R^2 results of 0.53 and 0.76, respectively, for SPEI-3 [116]. In contrast, the SVM algorithm applied to Combined Terrestrial Evapotranspiration Index (CTEI) resulted in an R^2 value of 0.82 and the lowest errors in terms of the root mean squared error (RMSE) (0.33) and Mean Absolute Error (MAE) (0.20) in the Ganga River basin [39] which agree with our results, and were much better in comparison to the other ML models they applied.

It is hard to determine a particular factor for the above results; it may be linked to the mutual causal relation between drought criteria and the SPEI. Drought criteria tend to represent the impact of rainfall shortages accumulated over the long-term rather than through short term episodes [117]. Also, RF was applied for short-term drought forecasting in the East Asia region using satellite-based climatic data [115]. This study showed that RF can capture the sudden change in drought conditions when a Madden–Julian oscillation is used as a feature for drought prediction. The researchers found that the overall average prediction accuracy is higher than the weather system's prediction accuracy, suggesting acceptable prediction results. Comparing models driven by remote sensing data only, and the combination of meteorological and remote sensing data, [114] stated that the SVM model driven by combined meteorological and remote sensing data performs better in the Guan Zhong area, China. For our second-best XGB scenario, our results were similar to the findings of [36] in Shaanxi, China, who concluded that the XGB model more precisely forecasted SPEI than the distributed lag nonlinear model (DLNM) and the artificial neural network (ANN), with R value ranges of 0.68–0.82 and 0.72–0.89 for SPEL-3 and SPEI-6, respectively.

CNN and LSTM models were also implemented to estimate the SPEIs using the seven climate input combination scenarios (Table 2), and the results are presented in Figure 8 and Table 4. For CNN-SPEI-3 models, the MSE, MAE, and R varied between 0.29 and 0.57, 0.49 and 0.71, and 0.57 and 0.82, respectively. Based on the NSE index, the performance of the CNN model is classified as a good fit for scenarios 6 and 7 and for scenarios 4 to 7 is classified

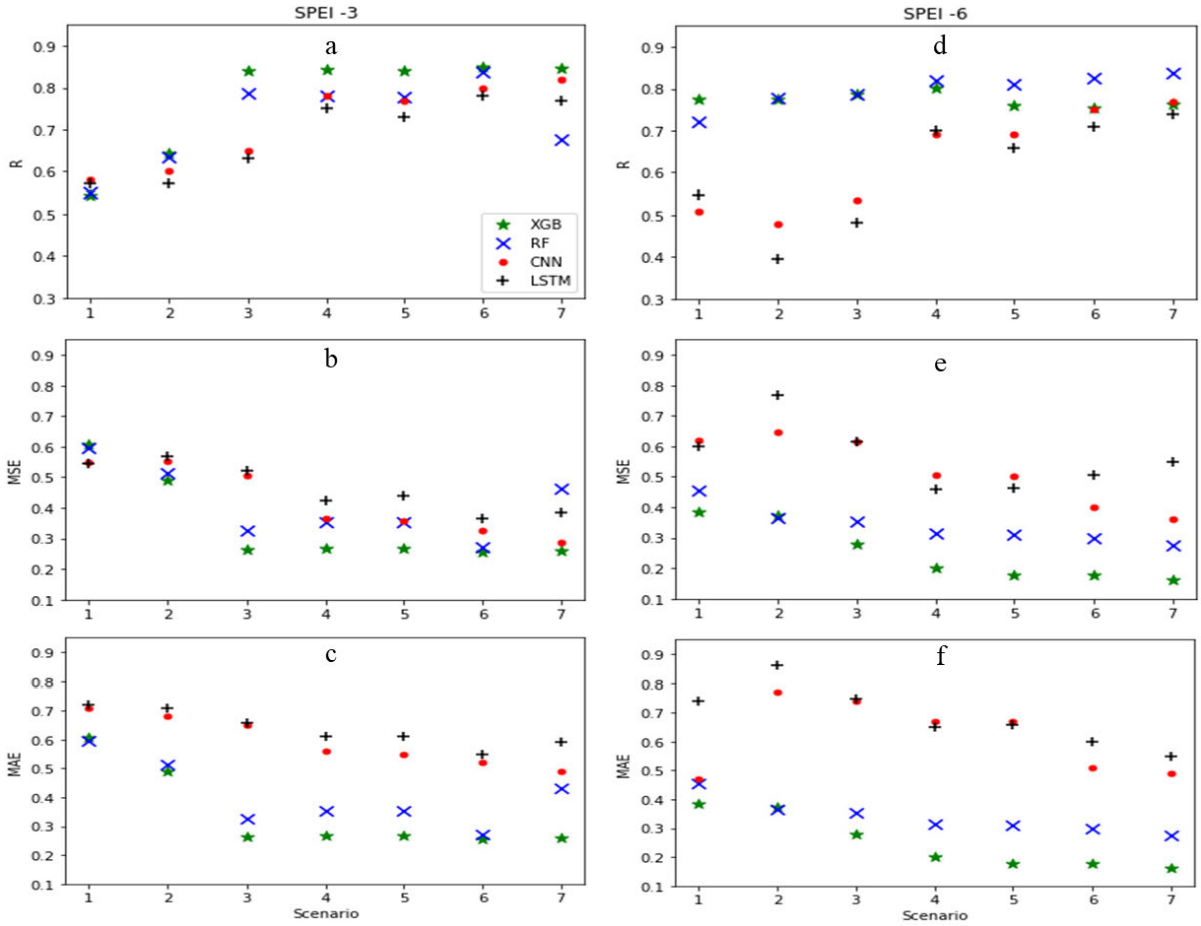

FIGURE 8. Performance statistics of the applied ML models for the 7 different climate variable combination scenarios.

as a satisfactory fit for both timescales. The CNN7 was the best scenario in comparison with the other CNN models as judged by MSE (0.29) and R (0.82). CNN6 exhibited the second-best performance in SPEI-3. The MSE, MAE, and R for LSTM6 ranged from 0.37 to 0.57, 0.49 to 0.72, and 0.57 to 0.78, respectively. The LSTM model with scenario 6 exhibited the best performance for estimating SPEI-3 as it had the lowest MSE (0.37), the lowest MAE (0.56) and the highest R (0.78). LSTM2 gave the worst result with the highest MSE of 0.56, which means that the combination of precipitation, average temperature, and maximum temperature was not sufficient for the estimation of SPEI-3 based on the LSTM model.

For SPEI-6 scenarios, the CNN and LSTM models exhibited the highest R values for scenario 7. CNN7 had a MSE of 0.36, MAE of 0.49, R of 0.77 and MBE of 0.08 compared to the other CNN models (Table 3). The results also indicated that CNN with scenarios 4, 5 and 6 could be good for estimating SPEI-6, and it is recommended when data are scarce. Moreover, CNN1 and CNN2 had the highest values of MSE which reached 0.77, and the lowest correlation ranging from 0.40 to 0.55. Also, LSTM 4 and 5 were good models for estimating SPEI-6, having the lowest MSE of 0.46, and R of 0.77 and 0.66, respectively. Our findings agree with [55] who applied the LSTM model to forecast SPEI-1 and SPEI-3 in New South Wales (Australia). Moreover, these results are close to the investigation of [57] who applied LSTM to forecast SPEI during the period of 1958-2014. Generally, the results of the SPEI-3 and SPEI-6 are close to each other for the four applied models due to less fluctuation in both timescales. This finding was similar to the investigation that applied WAANN and WANFIS models to predict the SPEI at the Langat River Basin for 1-, 3- and 6-months timescales. In terms of the 7 scenarios applied in this study, overall results of the first three scenarios were weak in comparison with the other four scenarios. The main reason is limited data (only precipitation and temperature) that are not able to achieve better accuracy of SPEI prediction. Moreover, the significant correlation of wind speed and relative humidity with SPEI is responsible for better prediction of SPEI in our study area.

C. COMPARISON OF THE CALCULATED AND SIMULATED SPEIs

Taylor diagrams were used to determine how much the SPEIs estimated by the ML models matched the calculated SPEIs (Fig. 9a, b). It showed the best combination of each model for each timescale. For the applied models, the best scenarios for the XGB were scenarios 5 and 7 for SPEI-3 and SPEI-6,

TABLE 4. Performance statistics of the XGB and RF model scenarios for the five zones.

Pattern	scenario	SPEI-3									
		R		NSE		MSE		MBE		MAE	
		XGB	RF	XGB	RF	XGB	RF	XGB	RF	XGB	RF
PC 1	4	0.91	0.90	0.79	0.80	0.17	0.18	0.01	0.02	0.35	0.36
	5	0.91	0.92	0.84	0.84	0.15	0.14	0.00	0.00	0.33	0.31
	6	0.94	0.94	0.88	0.87	0.11	0.11	-0.01	-0.01	0.26	0.27
PC 2	4	0.90	0.92	0.81	0.84	0.15	0.12	0.02	0.04	0.31	0.29
	5	0.92	0.92	0.84	0.85	0.13	0.12	0.02	0.03	0.30	0.28
	6	0.94	0.94	0.89	0.88	0.09	0.09	0.01	0.04	0.24	0.25
PC 3	4	0.92	0.91	0.85	0.82	0.15	0.17	0.00	0.02	0.33	0.35
	5	0.93	0.92	0.86	0.76	0.14	0.15	0.04	-0.03	0.31	0.33
	6	0.94	0.94	0.89	0.77	0.12	0.14	0.06	-0.01	0.28	0.31
PC 4	4	0.90	0.90	0.80	0.81	0.15	0.13	-0.04	0.00	0.32	0.31
	5	0.90	0.90	0.83	0.81	0.14	0.14	0.00	0.02	0.31	0.31
	6	0.92	0.91	0.86	0.82	0.14	0.15	0.04	0.03	0.30	0.32
PC 5	4	0.90	0.91	0.82	0.81	0.15	0.15	0.02	0.05	0.32	0.33
	5	0.92	0.93	0.83	0.83	0.14	0.14	0.06	0.05	0.30	0.31
	6	0.91	0.89	0.82	0.78	0.15	0.16	0.03	0.04	0.32	0.32
Average		0.92	0.92	0.84	0.82	0.14	0.14	0.02	0.02	0.31	0.31
		SPEI -6									
PC 1	4	0.90	0.92	0.82	0.84	0.17	0.15	0.00	-0.02	0.33	0.32
	5	0.92	0.92	0.84	0.84	0.16	0.15	0.02	0.03	0.33	0.33
	6	0.95	0.95	0.91	0.89	0.09	0.10	-0.01	0.02	0.25	0.26
PC 2	4	0.88	0.86	0.78	0.74	0.21	0.24	0.03	-0.01	0.36	0.39
	5	0.91	0.92	0.83	0.83	0.18	0.16	-0.02	0.01	0.34	0.33
	6	0.94	0.95	0.89	0.86	0.11	0.13	0.01	0.02	0.28	0.30
PC 3	4	0.95	0.92	0.80	0.83	0.11	0.18	-0.01	-0.01	0.27	0.33
	5	0.92	0.93	0.84	0.85	0.17	0.15	-0.01	0.00	0.33	0.31
	6	0.95	0.95	0.91	0.88	0.10	0.11	0.00	-0.01	0.24	0.27
PC 4	4	0.86	0.84	0.73	0.67	0.25	0.27	0.00	0.01	0.41	0.43
	5	0.89	0.87	0.77	0.72	0.22	0.23	-0.01	0.02	0.39	0.41
	6	0.92	0.92	0.83	0.81	0.15	0.16	0.02	0.03	0.33	0.33
PC 5	4	0.85	0.85	0.71	0.70	0.24	0.24	-0.08	-0.09	0.41	0.41
	5	0.87	0.85	0.75	0.72	0.21	0.19	-0.06	-0.05	0.39	0.37
	6	0.91	0.87	0.70	0.76	0.15	0.17	-0.02	-0.03	0.32	0.34
Average		0.91	0.90	0.81	0.80	0.17	0.18	-0.01	-0.01	0.33	0.34

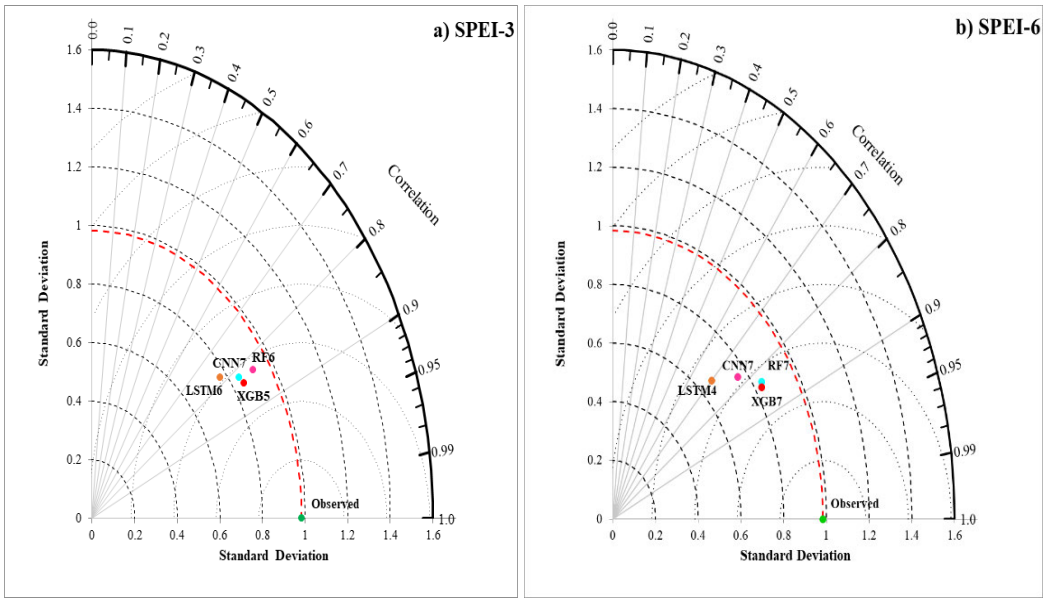

FIGURE 9. Taylor diagrams displaying a statistical comparison of applied models in the estimation of SPEI-3 (a) and SPEI-6 (b).

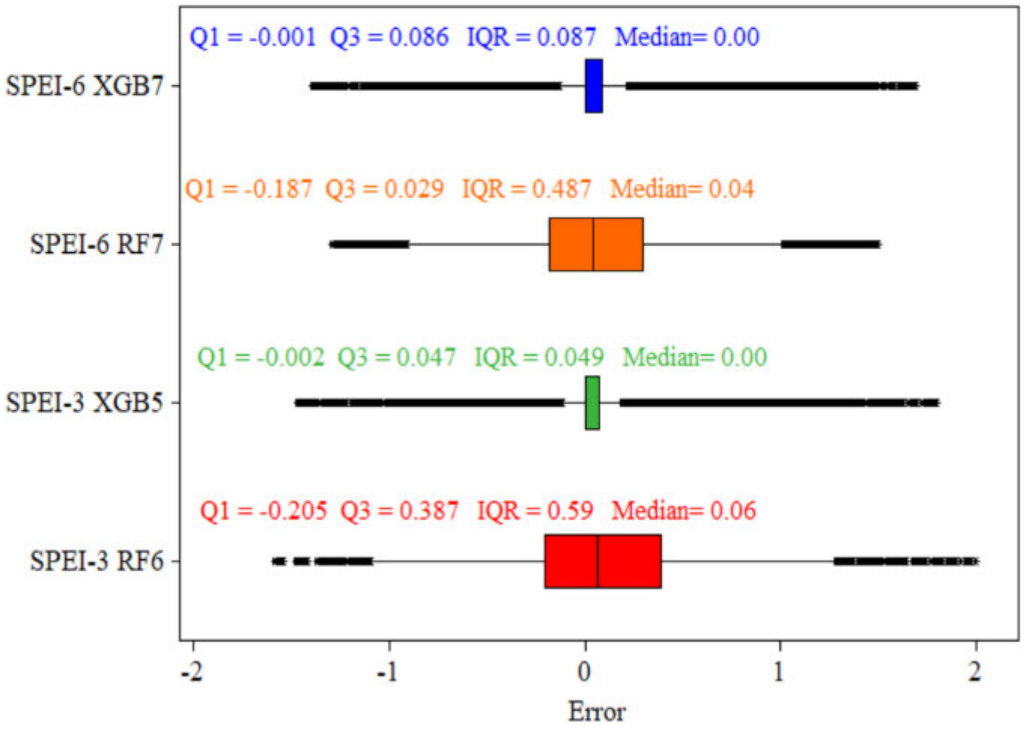

FIGURE 10. Boxplots showing the distribution of the estimation errors of the SPEIs in the test section for the two-best model-scenarios. Q1 is lower quartile of errors, Q3 is upper quartile of errors, and IQR is interquartile range for each model.

respectively, while the RF model was best with scenarios 6 and 7 for SPEI-3 and SPEI-6, respectively. For the DL models, CNN was superior for scenario 7 for both SPEIs, while for the LSTM scenario 6 was the best for SPEI-3 and scenario 4 for SPEI-6. Based on the Taylor diagram, the overall best models for SPEI-3 were XGB5 and RF6, while XGB7 and RF7 were the best models for SPEI-6. These model-scenario combinations generated the highest correlation, highest performance, and lowest deviations between the calculated and the forecasted SPEI values.

To determine the best model for the two SPEIs, boxplots were developed for the errors (residuals) estimated as the difference between the calculated and estimated values (Figure 10). The presented positive and negative estimation

errors indicate underestimation and overestimation, respectively. Therefore, XGB5 is the best model for SPEI-3 as it has the lowest error in comparison with RF6, and it is based on the main climatic variables (precipitation, temperature, wind speed and relative humidity). It performed better than RF6 for SPEI-3 (all input parameters except solar radiation) because it has the lower values of Q1, Q3 and IQR. Q3 (upper quartile) is more important than Q1 in error analysis because it contains 75% of the error. XGB5 for SPEI-3 has a smaller IQR compared with the best model-scenario for SPEI-6 (RF7), the difference being 0.438. Moreover, the IQR of XGB5 being the lowest among all models clearly indicated the superiority of the model. It needs to be underlined that the error distribution is around zero, and also the median line at the center of the rectangle shows normality of the error distribution for all four models (Figure 10).

Based on the above results which presented the best models as XGB and RF models for the whole area, but better performance was expected from the models. For this reason, the performance of the best models (i.e., RF and XGB) was evaluated separately in the five PCA regions/zones (at the local scale). Table 4 shows the performance statistics of each zone for the best models (XGB and RF). Input scenarios 4, 5, and 6 were used here because they achieved the best results in the previous stage evaluation. Moreover, these scenarios are the intermediate between using all input variables and the lowest inputs. For SPEI-3, NSE for all scenarios in the five zones was more than 0.75 which underlines the very good performance of the XGB and RF models according to [102]. Generally, the performance of the models is classified as very good for both XGB (NSE = 0.84) and RF (NSE = 0.82). Furthermore, the correlation coefficients exceed 0.89 for all of the models in all zones, especially in zones 1, 2 and 3 where NSE reached 0.94 for the XGB model. For the five zones, the average R was 0.92 for both XGB and RF models. The MSE ranged from 0.09 to 0.18 for the five zones with an average value of 0.14. For SPEI-6, and based on NSE, all scenarios in the five zones are classified as very good fits for XGB and RF models except for some zones that are classified as good fits. The highest R was in zone 1 which reached 0.95 followed by zones 2 and 3. MBE values for both scales show low values that confirm the ability of the models to estimate SPEI in the study areas. Table 4 shows that the RF and XGB models have promising results in estimating SPEI. However, the performance of the models will increase significantly if used separately in each zone. As the PCA approach showed, this region has 5 parts with different climatic characteristics, and this can affect the performance of comprehensive models covering the whole region.

V. CONCLUSION

Drought events identification using SPEI is carried out in this study. More than 30% of the studied area has experienced extreme and very extreme droughts during the period of 2010-2019. Moreover, during the period of 2015-2019, the spatial distribution of the drought covered 80-90% of the area under investigation, except for some northern and southwestern parts for SPEI-3 and SPEI-6. Our results clearly indicate a significant negative trend for SPEI-3 and SPEI-6 over the study period. In this regard, an adaptation strategy on a national scale should be implemented to minimize the direct impact of the drought cycles.

Further, four machine learning models (XGB, RF, LSTM and CNN) were developed to estimate the SPEI on 3- and 6-month timescales by using various combinations of climate variables as predictors. The XGB5 (relying on temperature, wind, RH) was found to be a superior model for SPEI-3 and XGB7 (relying on temperature, wind, RH, Sunshine, SR) was superior for SPEI-6 prediction. For the DL models, LSTM with scenarios 6 and 4 were the best for SPEI-3 and SPEI-6, respectively. For SPEI-3, NSE for all scenarios in the five zones was more than 0.75 which underlines the very good performance of the XGB and RF models. The RF and XGB models have showed promising results in estimating SPEI and produce a more practical model at the regional scale, and not only for a specific area or station. However, the performance of the models increases significantly when used separately in each zone. Thus, this investigation can present a pioneering modeling strategy that could lead to improvement of efforts to address the deficiencies in drought prediction, sustainable water-use, and the management of water supply systems. The results show that the different climatic characteristics in each station are the major reasons for the weak results in low-input scenarios. However, the low-input combinations achieved good results in the prediction of SPEI on the local scale. In future studies, more predictors need to be considered to better understand the mechanism of drought such as atmospheric circulation, natural variability, and anthropogenic activities. Improving drought prediction methods is one important approach to facilitate a transition to more sustainable water resource management in the context of drought by focusing more on Multi-step Time Series Forecasting of drought using a hybrid deep learning.

AUTHOR CONTRIBUTIONS

AM and HH collected and analyzed the research data. AM and MJ designed and applied the machine learning models of the research. AM, AE and KA wrote the original manuscript and provided suggestions regarding data analyses. AM and KA generated the figures in the main text. AM, HG-A, YGA, and JR-C revised and provided suggestions to improve the contents and structure of the paper. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

[1] W. Jiang, L. Wang, M. Zhang, R. Yao, X. Chen, X. Gui, J. Sun, and Q. Cao, "Analysis of drought events and their impacts on vegetation productivity based on the integrated surface drought index in the Hanjiang River Basin, China," *Atmos. Res.*, vol. 254, Jun. 2021, Art. no. 105536.

[2] T. Y. Gan, "Possible climate change/variability and human impacts, vulnerability of drought-prone regions, water resources and capacity building for Africa," *Hydrol. Sci. J.*, vol. 61, no. 7, pp. 1209–1226, 2016.

[3] R. Zhong, X. Chen, C. Lai, Z. Wang, Y. Lian, H. Yu, and X. Wu, "Drought monitoring utility of satellite-based precipitation products across mainland China," *J. Hydrol.*, vol. 568, pp. 343–359, Jan. 2019.
[4] S. M. Vicente-Serrano, G. Van der Schrier, S. Beguería, C. Azorin-Molina, and J.-I. Lopez-Moreno, "Contribution of precipitation and reference evapotranspiration to drought indices under different climates," *J. Hydrol.*, vol. 526, pp. 42–54, Jul. 2015.
[5] N. Khan, "Prediction of droughts over Pakistan using machine learning algorithms," *Adv. Water Resour.*, vol. 139, May 2020, Art. no. 103562.
[6] M. C. Anderson, C. A. Zolin, P. C. Sentelhas, C. R. Hain, K. Semmens, M. T. Yilmaz, F. Gao, J. A. Otkin, and R. Tetrault, "The evaporative stress index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts," *Remote Sens. Environ.*, vol. 174, pp. 82–99, Mar. 2016.
[7] D. E. Akyuz, M. Bayazit, and B. Onoz, "Markov chain models for hydrological drought characteristics," *J. Hydrometeorol.*, vol. 13, no. 1, pp. 298–309, Feb. 2012.
[8] N. Khan, S. Shahid, E.-S. Chung, S. Kim, and R. Ali, "Influence of surface water bodies on the land surface temperature of Bangladesh," *Sustainability*, vol. 11, no. 23, p. 6754, Nov. 2019.
[9] P. Paneque, "Drought management strategies in Spain," *Water*, vol. 7, no. 12, pp. 6689–6701, Nov. 2015.
[10] L. Xu, N. Chen, X. Zhang, and Z. Chen, "An evaluation of statistical, NMME and hybrid models for drought prediction in China," *J. Hydrol.*, vol. 566, pp. 235–249, Nov. 2018.
[11] J. Qiu, "China drought highlights future climate threats: Yunnan's worst drought for many years has been exacerbated by destruction of forest cover and a history of poor water management," *Nature*, vol. 465, no. 7295, pp. 142–144, 2010.
[12] K. Xu, D. Yang, H. Yang, Z. Li, Y. Qin, and Y. Shen, "Spatio-temporal variation of drought in China during 1961–2012: A climatic perspective," *J. Hydrol.*, vol. 526, pp. 253–264, Jul. 2015.
[13] I. Bordi and A. Sutera, "Drought monitoring and forecasting at large scale," in *Methods and Tools for Drought Analysis and Management*. Dordrecht, The Netherlands: Springer, 2007, pp. 3–27.
[14] Y. W. Soh, C. H. Koo, Y. F. Huang, and K. F. Fung, "Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, malaysia," *Comput. Electron. Agricult.*, vol. 144, pp. 164–173, Jan. 2018.
[15] K. Beven, "A manifesto for the equifinality thesis," *J. Hydrol.*, vol. 320, nos. 1–2, pp. 18–36, Mar. 2006.
[16] J. F. Adamowski, "Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis," *J. Hydrol.*, vol. 353, nos. 3–4, pp. 247–266, May 2008.
[17] A. Elbeltagi, M. R. Aslam, A. Malik, B. Mehdinejadiani, A. Srivastava, A. S. Bhatia, and J. Deng, "The impact of climate changes on the water footprint of wheat and maize production in the Nile Delta, Egypt," *Sci. Total Environ.*, vol. 743, Nov. 2020, Art. no. 140770.
[18] M. K. Abdel-Fattah and A. A. I. M. Abdo, "Application of neural network and time series modeling to study the suitability of drain water quality for irrigation: A case study from Egypt," *Environ. Sci. Pollut. Res.*, vol. 28, no. 1, pp. 1–17, 2020.
[19] S. Maroufpoor, E. Maroufpoor, O. Bozorg-Haddad, J. Shiri, and Z. M. Yaseen, "Soil moisture simulation using hybrid artificial intelligent model: Hybridization of adaptive neuro fuzzy inference system with grey wolf optimizer algorithm," *J. Hydrol.*, vol. 575, pp. 544–556, Aug. 2019.
[20] P. Ganguli and M. J. Reddy, "Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach," *Hydrol. Processes*, vol. 28, no. 19, pp. 4989–5009, Sep. 2014.
[21] Y. Feng, N. Cui, Y. Chen, D. Gong, and X. Hu, "Development of data-driven models for prediction of daily global horizontal irradiance in northwest China," *J. Cleaner Prod.*, vol. 223, pp. 136–146, Jun. 2019.
[22] F. Granata, "Evapotranspiration evaluation models based on machine learning algorithms—A comparative study," *Agricult. Water Manage.*, vol. 217, pp. 303–315, May 2019.
[23] S. Mouatadid, N. Raj, R. C. Deo, and J. F. Adamowski, "Input selection and data-driven model performance optimization to predict the standardized precipitation and evaporation index in a drought-prone region," *Atmos. Res.*, vol. 212, pp. 130–149, Nov. 2018.
[24] W. Pozzi *et al.*, "Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting," *Bull. Amer. Meteorol. Soc.*, vol. 94, no. 6, pp. 776–785, Jun. 2013.
[25] S. Shahid, "Rainfall variability and the trends of wet and dry periods in Bangladesh," *Int. J. Climatol.*, vol. 30, no. 15, pp. 2299–2313, Dec. 2010.
[26] A. K. Mishra and V. P. Singh, "Drought modeling—A review," *J. Hydrol.*, vol. 403, nos. 1–2, pp. 157–175, 2011.
[27] A. Belayneh, J. Adamowski, B. Khalil, and B. Ozga-Zielinski, "Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models," *J. Hydrol.*, vol. 508, pp. 418–429, Jan. 2014.
[28] M. Masinde, "Artificial neural networks models for predicting effective drought index: Factoring effects of rainfall variability," *Mitigation Adaptation Strategies Global Change*, vol. 19, no. 8, pp. 1139–1162, Dec. 2014.
[29] V. H. Quej, J. Almorox, J. A. Arnaldo, and L. Saito, "ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment," *J. Atmos. Solar-Terr. Phys.*, vol. 155, pp. 62–70, Mar. 2017.
[30] J. Fan, L. Wu, F. Zhang, H. Cai, X. Wang, X. Lu, and Y. Xiang, "Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature," *Renew. Sustain. Energy Rev.*, vol. 94, pp. 732–747, Oct. 2018.
[31] R. Kumar, R. K. Aggarwal, and J. D. Sharma, "Comparison of regression and artificial neural network models for estimation of global solar radiations," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1294–1299, Dec. 2015.
[32] A. Moreno, M. A. Gilabert, and B. Martínez, "Mapping daily global solar irradiation over spain: A comparative study of selected approaches," *Sol. Energy*, vol. 85, no. 9, pp. 2072–2084, Sep. 2011.
[33] S. H. H. Nazhad, M. M. Lotfinejad, M. Danesh, R. Ul Amin, and S. Shamshirband, "A comparison of the performance of some extreme learning machine empirical models for predicting daily horizontal diffuse solar radiation in a region of southern Iran," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 6894–6909, Dec. 2017.
[34] S. Ghimire, R. C. Deo, N. J. Downs, and N. Raj, "Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia," *J. Cleaner Prod.*, vol. 216, pp. 288–310, Apr. 2019.
[35] Y. Radhika and M. Shashi, "Atmospheric temperature prediction using support vector machines," *Int. J. Comput. theory Eng.*, vol. 1, no. 1, p. 55, 2009.
[36] R. Zhang, Z.-Y. Chen, L.-J. Xu, and C.-Q. Ou, "Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China," *Sci. Total Environ.*, vol. 665, pp. 338–346, May 2019.
[37] E. E. Başakın, Ö. Ekmekcioğlu, and M. Ozger, "Drought analysis with machine learning methods," *Pamukkale Univ. J. Eng. Sci.*, vol. 25, no. 8, pp. 985–991, 2019.
[38] A. N. Shahbazi, "Seasonal meteorological drought prediction using support vector machine," *World Appl. Sci. J.*, vol. 13, no. 6, pp. 1387–1397, 2011.
[39] A. Elbeltagi, N. Kumari, J. K. Dharpure, A. Mokhtar, K. Alsafadi, M. Kumar, B. Mehdinejadiani, H. R. Etedali, Y. Brouziyne, A. R. M. T. Islam, and A. Kuriqi, "Prediction of combined terrestrial evapotranspiration index (CTEI) over large river basin based on machine learning approaches," *Water*, vol. 13, no. 4, p. 547, Feb. 2021.
[40] P. J. Starks, J. L. Steiner, J. P. S. Neel, K. E. Turner, B. K. Northup, P. H. Gowda, and M. A. Brown, "Assessment of the standardized precipitation and evaporation index (SPEI) as a potential management tool for grasslands," *Agronomy*, vol. 9, no. 5, p. 235, May 2019.
[41] S. Mouatadid, R. C. Deo, and J. F. Adamowski, "Prediction of SPEI using MLR and ANN: A case study for Wilsons Promontory Station in Victoria," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 318–324.
[42] R. C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia," *Atmos. Res.*, vol. 153, pp. 512–525, Feb. 2015.
[43] S. Papadopoulos, E. Azar, W.-L. Woon, and C. E. Kontokosta, "Evaluation of tree-based ensemble learning algorithms for building energy performance estimation," *J. Building Perform. Simul.*, vol. 11, no. 3, pp. 322–332, May 2018.
[44] Y. Feng, N. Cui, Q. Zhang, L. Zhao, and D. Gong, "Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in north China plain," *Int. J. Hydrogen Energy*, vol. 42, no. 21, pp. 14418–14428, May 2017.

[45] M. A. Hassan, A. Khalil, S. Kaseb, and M. A. Kassem, "Exploring the potential of tree-based ensemble methods in solar radiation modeling," *Appl. Energy*, vol. 203, pp. 897–916, Oct. 2017.

[46] L. B. Ferreira, F. F. da Cunha, R. A. de Oliveira, and E. I. F. Filho, "Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—A new approach," *J. Hydrol.*, vol. 572, pp. 556–570, May 2019.

[47] M. K. Saggi and S. Jain, "Reference evapotranspiration estimation and modeling of the Punjab northern India using deep learning," *Comput. Electron. Agricult.*, vol. 156, pp. 387–398, Jan. 2019.

[48] Q. Guo, Y. Li, Y. Song, D. Wang, and W. Chen, "Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2044–2053, Mar. 2020.

[49] T. Lee, J.-Y. Shin, J.-S. Kim, and V. P. Singh, "Stochastic simulation on reproducing long-term memory of hydroclimatological variables using deep learning model," *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124540.

[50] H. Son and C. Kim, "A deep learning approach to forecasting monthly demand for Residential–Sector electricity," *Sustainability*, vol. 12, no. 8, p. 3103, Apr. 2020.

[51] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, and Y.-S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts," *J. Cleaner Prod.*, vol. 209, pp. 134–145, Feb. 2019.

[52] R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model," *Stochastic Environ. Res. Risk Assessment*, vol. 34, pp. 415–433, 2020, doi: 10.1007/s00477-020-01776.

[53] A. Sayeed, Y. Choi, E. Eslami, Y. Lops, A. Roy, and J. Jung, "Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance," *Neural Netw.*, vol. 121, pp. 396–408, Jan. 2020.

[54] R. Shen, A. Huang, B. Li, and J. Guo, "Construction of a drought monitoring model using deep learning based on multi-source remote sensing data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 79, pp. 48–57, Jul. 2019.

[55] A. Dikshit, B. Pradhan, and A. Huete, "An improved SPEI drought forecasting approach using the long short-term memory neural network," *J. Environ. Manage.*, vol. 283, Apr. 2021, Art. no. 111979.

[56] A. Dikshit, B. Pradhan, and A. M. Alamri, "Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model," *Sci. Total Environ.*, vol. 755, Feb. 2021, Art. no. 142638.

[57] S. Poornima and M. Pushpalatha, "Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network," *Soft Comput.*, vol. 23, no. 18, pp. 8399–8412, Sep. 2019.

[58] J. Peng, Z. Liu, Y. Liu, J. Wu, and Y. Han, "Trend analysis of vegetation dynamics in Qinghai–Tibet plateau using hurst exponent," *Ecol. Indicators*, vol. 14, no. 1, pp. 28–39, Mar. 2012.

[59] Z. Yuke, "Characterizing the spatio-temporal dynamics and variability in climate extremes over the Tibetan plateau during 1960–2012," *J. Resour. Ecol.*, vol. 10, no. 4, pp. 397–414, 2019.

[60] Y. H. Yang, J. Y. Fang, Y. D. Pan, and C. J. Ji, "Aboveground biomass in tibetan grasslands," *J. Arid Environ.*, vol. 73, no. 1, pp. 91–95, Jan. 2009.

[61] Y. Xu, V. Ramanathan, and W. M. Washington, "Observed high-altitude warming and snow cover retreat over tibet and the Himalayas enhanced by black carbon aerosols," *Atmos. Chem. Phys.*, vol. 16, no. 3, pp. 1303–1315, Feb. 2016.

[62] D. Yihui and J. C. L. Chan, "The East Asian summer monsoon: An overview," *Meteorol. Atmos. Phys.*, vol. 89, nos. 1–4, pp. 117–142, Jun. 2005.

[63] R. Zhang, A. Sumi, and M. Kimoto, "A diagnostic study of the impact of el Niño on the precipitation in China," *Adv. Atmos. Sci.*, vol. 16, no. 2, pp. 229–241, May 1999.

[64] D. Wang, H. He, Q. Gao, W. He, C. Zhao, H. Yin, and Q. Liu, "Effects of short-term N addition on soil C fluxes in alpine Sibiraea angustata scrub on the eastern margin of the Qinghai-Tibetan Plateau," *Agricult. Forest Meteorol.*, vol. 247, pp. 151–158, Dec. 2017.

[65] T. Yao, L. Thompson, W. Yang, W. Yu, Y. Gao, X. Guo, X. Yang, K. Duan, H. Zhao, B. Xu, J. Pu, A. Lu, Y. Xiang, D. B. Kattel, and D. Joswiak, "Different glacier status with atmospheric circulations in tibetan plateau and surroundings," *Nature Climate Change*, vol. 2, no. 9, pp. 663–667, Sep. 2012.

[66] A. Mokhtar, H. He, K. Alsafadi, Y. Li, H. Zhao, S. Keo, C. Bai, M. Abuarab, C. Zhang, K. Elbagoury, J. Wang, and Q. He, "Evapotranspiration as a response to climate variability and ecosystem changes in southwest, China," *Environ. Earth Sci.*, vol. 79, no. 12, p. 312, Jun. 2020.

[67] A. Mokhtar, H. He, H. Zhao, S. Keo, C. Bai, C. Zhang, Y. Ma, A. Ibrahim, Y. Li, F. Li, W. He, A. I. Abdo, and J. Zhou, "Risks to water resources and development of a management strategy in the river basins of the Hengduan Mountains, Southwest China," *Environ. Sci., Water Res. Technol.*, vol. 6, no. 3, pp. 656–678, 2020.

[68] R. G. Allen, "Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study," *J. Hydrol.*, vol. 229, nos. 1–2, pp. 27–41, Mar. 2000.

[69] Z.-X. Fan and A. Thomas, "Spatiotemporal variability of reference evapotranspiration and its contributing climatic factors in Yunnan Province, SW China, 1961–2004," *Climatic Change*, vol. 116, no. 2, pp. 309–325, Jan. 2013.

[70] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index," *J. Climate*, vol. 23, no. 7, pp. 1696–1718, Apr. 2010.

[71] Y. Yihdego, B. Vaheddoost, and R. A. Al-Weshah, "Drought indices and indicators revisited," *Arabian J. Geosci.*, vol. 12, no. 3, p. 69, Feb. 2019.

[72] J. Li, Y. Yue, H. Pan, and X. Y. Ye, "Variation rules of meteorological drought in China during 1961–2010 based on SPEI and intensity analysis," *J. Catastrophol.*, vol. 29, no. 4, pp. 176–182, 2014.

[73] B. Ming, Y.-Q. Guo, H.-B. Tao, G.-Z. Liu, S.-K. Li, and P. Wang, "SPEIPM-based research on drought impact on maize yield in north China plain," *J. Integrative Agricult.*, vol. 14, no. 4, pp. 660–669, Apr. 2015.

[74] W. Jiang, L. Wang, L. Feng, M. Zhang, and R. Yao, "Drought characteristics and its impact on changes in surface vegetation from 1981 to 2015 in the Yangtze River Basin, China," *Int. J. Climatol.*, vol. 40, no. 7, pp. 3380–3397, Jun. 2020.

[75] W. Wang, Y. Zhu, R. Xu, and J. Liu, "Drought severity change in China during 1961–2012 indicated by SPI and SPEI," *Natural Hazards*, vol. 75, no. 3, pp. 2437–2451, Feb. 2015.

[76] J. Park, Y.-J. Lim, B.-J. Kim, and J. H. Sung, "Appraisal of drought characteristics of representative drought indices using meteorological variables," *KSCE J. Civil Eng.*, vol. 22, no. 5, pp. 2002–2009, May 2018.

[77] C. T. Agnew, "Using the SPI to identify drought," *Drought Netw. News*, vol. 12, no. 1, 2000.

[78] A. M. El Kenawy, A. Al Buloshi, T. Al-Awadhi, N. Al Nasiri, F. Navarro-Serrano, S. Alhatrushi, S. M. Robaa, F. Domínguez-Castro, M. F. McCabe, P.-M. Schuwerack, J. I. López-Moreno, and S. M. Vicente-Serrano, "Evidence for intensification of meteorological droughts in oman over the past four decades," *Atmos. Res.*, vol. 246, Dec. 2020, Art. no. 105126.

[79] K. Alsafadi, S. A. Mohammed, B. Ayugi, M. Sharaf, and E. Harsányi, "Spatial–temporal evolution of drought characteristics over Hungary between 1961 and 2010," *Pure Appl. Geophys.*, vol. 177, no. 8, pp. 3961–3978, Aug. 2020.

[80] B. Lloyd-Hughes, "A spatio-temporal structure-based approach to drought characterisation," *Int. J. Climatol.*, vol. 32, no. 3, pp. 406–418, Mar. 2012.

[81] C. Tan, J. Yang, and M. Li, "Temporal-spatial variation of drought indicated by SPI and SPEI in Ningxia Hui Autonomous Region, China," *Atmosphere*, vol. 6, no. 10, pp. 1399–1421, Sep. 2015.

[82] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Proc. 8th Conf. Appl. Climatol.*, Boston, MA, USA, 1993, pp. 179–183.

[83] J. Spinoni, G. Naumann, H. Carrao, P. Barbosa, and J. Vogt, "World drought frequency, duration, and severity for 1951–2010," *Int. J. Climatol.*, vol. 34, no. 8, pp. 2792–2804, Jun. 2014.

[84] H. Guo, A. Bao, T. Liu, F. Ndayisaba, D. He, A. Kurban, and P. De Maeyer, "Meteorological drought analysis in the lower mekong basin using satellite-based long-term CHIRPS product," *Sustainability*, vol. 9, no. 6, p. 901, May 2017.

[85] H. Guo, A. Bao, T. Liu, G. Jiapaer, F. Ndayisaba, L. Jiang, A. Kurban, and P. De Maeyer, "Spatial and temporal characteristics of droughts in Central Asia during 1966–2015," *Sci. Total Environ.*, vol. 624, pp. 1523–1538, May 2018.

[86] L. Yan-Jun, Z. Xiao-Dong, L. Fan, and M. Jing, "Analysis of drought evolvement characteristics based on standardized precipitation index in the Huaihe River basin," *Procedia Eng.*, vol. 28, pp. 434–437, Jan. 2012.

[87] S. M. Vicente-Serrano and J. I. López-Moreno, "The influence of atmospheric circulation at different spatial scales on winter drought variability through a semi-arid climatic gradient in northeast Spain," *Int. J. Climatol.*, vol. 26, no. 11, pp. 1427–1453, 2006.

[88] S. Mohammed, K. Alsafadi, T. Al-Awadhi, Y. Sherief, E. Harsanyie, and A. M. El Kenawy, "Space and time variability of meteorological drought in Syria," *Acta Geophysica*, vol. 68, no. 6, pp. 1877–1898, Dec. 2020.
[89] H. F. Kaiser, "Coefficient alpha for a principal component and the Kaiser-Guttman rule," *Psychol. Rep.*, vol. 68, no. 3, pp. 855–858, Jun. 1991.
[90] M. Portela, "Comprehensive characterization of droughts in Slovakia," *Int. J. Environ. Sci. Develop.*, vol. 8, no. 1, p. 25, 2017.
[91] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[92] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[93] L. B. Ferreira and F. F. da Cunha, "Multi-step ahead forecasting of daily reference evapotranspiration using deep learning," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105728.
[94] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[95] Q. Wu and H. Lin, "Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network," *Sustain. Cities Soc.*, vol. 50, Oct. 2019, Art. no. 101657.
[96] S. Zhu, B. Hrnjica, M. Ptak, A. Choiński, and B. Sivakumar, "Forecasting of water level in multiple temperate lakes using machine learning models," *J. Hydrol.*, vol. 585, Jun. 2020, Art. no. 124819.
[97] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[98] C. Tian, "A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network," *Energies*, vol. 11, no. 12, p. 3493, 2018.
[99] L. B. Ferreira and F. F. da Cunha, "New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning," *Agricult. Water Manage.*, vol. 234, May 2020, Art. no. 106113.
[100] R. Zuo, Y. Xiong, J. Wang, and E. J. M. Carranza, "Deep learning and its application in geochemical mapping," *Earth-Sci. Rev.*, vol. 192, pp. 1–14, May 2019.
[101] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Syst. Appl.*, vol. 129, pp. 273–285, Sep. 2019.
[102] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Trans. ASABE*, vol. 50, no. 3, pp. 885–900, 2007.
[103] Z. Yan, Y. Li, H. Wu, K. Zhang, Y. Hao, J. Wang, X. Zhang, L. Yan, and X. Kang, "Different responses of soil hydrolases and oxidases to extreme drought in an alpine peatland on the Qinghai-Tibet Plateau, China," *Eur. J. Soil Biol.*, vol. 99, Jul. 2020, Art. no. 103195.
[104] S. Wang, H. Duan, and J. Feng, "Drought events and its influence in autumn of 2014 in China," *J. Arid Meteorol.*, vol. 32, pp. 1031–1039, 2014.
[105] Z. Yu, F. Jianying, W. Zhilan, and W. Supin, "Drought events and its influence in autumn of 2018 in China," *J. Arid Meteorol.*, vol. 36, no. 6, p. 1052, 2018.
[106] Z. Niu, L. Wang, L. Fang, J. Li, and R. Yao, "Spatiotemporal variations in monthly relative humidity in China based on observations and CMIP5 models," *Int. J. Climatol.*, vol. 40, no. 15, pp. 6382–6395, Dec. 2020.
[107] H. Duan, S. Wang, and J. Feng, "Drought events and its influence in 2014 in China," *J. Arid Meteorol.*, vol. 33, pp. 349–360, 2015.
[108] Y. Li, H. Xu, and D. Liu, "Features of the extremely severe drought in the east of southwest China and anomalies of atmospheric circulation in summer 2006," *Acta Meteorol. Sinica*, vol. 25, no. 2, pp. 176–187, Apr. 2011.
[109] G. Leng, Q. Tang, and S. Rayburg, "Climate change impacts on meteorological, agricultural and hydrological droughts in China," *Global Planet. Change*, vol. 126, pp. 23–34, Mar. 2015.
[110] A. Wang, D. P. Lettenmaier, and J. Sheffield, "Soil moisture drought in China, 1950-2006," *J. Climate*, vol. 24, no. 13, pp. 3257–3271, 2011.
[111] J. Yang, D. Gong, W. Wang, M. Hu, and R. Mao, "Extreme drought event of 2009/2010 over southwestern China," *Meteorol. Atmos. Phys.*, vol. 115, nos. 3–4, pp. 173–184, Feb. 2012.
[112] E. Lu, Y. Luo, R. Zhang, Q. Wu, and L. Liu, "Regional atmospheric anomalies responsible for the 2009–2010 severe drought in China," *J. Geophys. Res.*, vol. 116, 2011, Art. no. D21114, doi: 10.1029/2011JD015706.
[113] X. X. Lu and Y. L. X. Siyue, "Dam not sole cause of Chinese drought," *Nature*, vol. 475, no. 7355, p. 174, 2011.
[114] J. Li, S. Zhang, L. Huang, T. Zhang, and P. Feng, "Drought prediction models driven by meteorological and remote sensing data in Guanzhong area, China," *Hydrol. Res.*, vol. 51, no. 5, pp. 942–958, Oct. 2020.
[115] S. Park, J. Im, E. Jang, and J. Rhee, "Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions," *Agricult. Forest Meteorol.*, vol. 216, pp. 157–169, Jan. 2016.
[116] A. Dikshit, B. Pradhan, and A. M. Alamri, "Short-term spatio-temporal drought forecasting using random forests model at New South Wales, Australia," *Appl. Sci.*, vol. 10, no. 12, p. 4254, Jun. 2020.
[117] U. Gessner, V. Naeimi, I. Klein, C. Kuenzer, D. Klein, and S. Dech, "The relationship between precipitation anomalies and satellite-derived vegetation activity in central Asia," *Global Planet. Change*, vol. 110, pp. 74–87, Nov. 2013.

ALI MOKHTAR received the B.Sc. and M.Sc. degrees in agriculture from the University of Cairo, Egypt, in 2012 and 2016, respectively, and the Ph.D. degree from Northwest A&F University, China, in 2020. He currently works with the Department of Agricultural Engineering, Cairo University. He does research in agricultural water resources management and climate change "Hydrological science." The current project is "Water resources management, ecosystem changes, climate change, and artificial intelligence." He is working in mapping and developing many models for analysis of the distribution of climatic elements and has applied that in hydrological science, water resources, climate research, water quality indexes, heavy metals exposure risks via water pathways, and pollution indices approaches along with the applications of artificial intelligence models in agricultural water dynamics.

MOHAMMADNABI JALALI is currently pursuing the Ph.D in water resources engineering from the Department of Water Engineering, University of Tehran, Iran. His specialized field of interest is climate change, water diplomacy, the use of intelligent models in water resources, and social hydrology.

HONGMING HE received the Ph.D. degree from the University of Chinese Academy of Sciences, China, in 2006. He is a Professor with the Institute of Soil and Water Conservation, Chinese Academy of Science, and the School of Geographic Sciences, East China Normal University, China. He does research in numerical simulation of environment system, including remote sensing, hydrology, erosion and fluvial geomorphology, and paleoclimatology. His current project is "Biodiversity and geospatial modeling."

NADHIR AL-ANSARI received the Ph.D. degree in water resources engineering from the University of Dundee, in 1976. He is currently a Professor with Luleå University of Technology, Sweden. He previously worked with Baghdad University, from 1976 to 1995, and with Al-Bayt University, Jordan, from 1995 to 2007. He has served several academic administrative posts, including the dean and the head of Department. His publications include more than 673 articles in journals and 21 books. He has executed more than 60 major research projects in Iraq, Jordan, and U.K.

AHMED ELBELTAGI currently works with the Department of Agricultural Engineering, Mansoura University. He does research in agricultural water management. The current project he is working on is "Applications of artificial systems in agricultural water dynamics." He currently serves as an Editorial Board Member for the *Journal of Agricultural Studies*, *American Journal of Agricultural and Biological Sciences*, *Agricultural Water Management*, and *Water Resources Management*.

KARAM ALSAFADI received the Ph.D. degree in physical geography from Alexandria University, Egypt, in 2020. He is currently a Postdoctoral Fellow with the Nanjing University of Information Science and Technology (NUIST). His research explores the spatial-temporal interactions between climate biosphere at a range of scales, especially in climatic risks, extreme events like drought, and climate change impacts on plant disease outbreaks in Mediterranean region. He also works on spatial modeling of observed climatic parameters for ecological management and long-term sustainability. His main research interests include climatology and agroclimatology, GIS, and remote sensing applications.

HAZEM GHASSAN ABDO received the B.Sc. and M.Sc. degrees in physical geography from Tishreen University, Lattakia, Syria, in 2013 and 2017, respectively, and the Ph.D. degree in earth system modeling (applied geomorphology) from Damascus University, Syria, in 2020. He is currently a Lecturer in the field of physical geography in the Geography Department, College of Arts and Humanities, Tartous University. His main research interest includes the use of remote sensing and geographic information systems for applications in the terrestrial environment to the understanding of earth system processes, in the context of environmental and geomorphological dynamic.

SAAD SH. SAMMEN received the master's degree from the University of Technology, Iraq, in 2009, and the Ph.D. degree from the University Putra Malaysia (UPM), Malaysia, in 2018. He is currently a Senior Lecturer and a Senior Researcher in civil engineering. He has published many articles in international journals. His current research interests include hydraulic engineering, water resources engineering, hydrological processes modeling, flood modeling, and climate change. In addition, he has an expertise in machine learning.

YEBOAH GYASI-AGYEI received the B.Sc. degree (Hons.) in civil engineering from the Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in 1984, and the M.Sc. and Ph.D. degrees in hydrology and water resources engineering from Free University Brussels, Belgium, in 1989 and 1993, respectively. He retired as an Associate Professor from CQ University, Australia, in 2020. He is currently an Adjunct Associate Professor with the School of Engineering and Built Environment, Griffith University, Australia. His research interests include catchment hydrology, spatio-temporal rainfall modeling, rainfall-runoff-erosion modeling, erosion control of steep engineered structures, and drip irrigation.

JESÚS RODRIGO-COMINO is currently working as a Junior Researcher/Postdoc with University of Trier, Germany, managing the European INTERREG-Project: Smart Light-HUB. He received a Grant from the COST-Action: CA18135 Firelinks at Valencia University. He is also the Chief Editor for Soil Section of *Air, Soil and Water Research* (SAGE).