



# Drought index downscaling using AI-based ensemble technique and satellite data

Nazanin Behfar<sup>1,2</sup> · Elnaz Sharghi<sup>2</sup> · Vahid Nourani<sup>2,3,4</sup> · Martijn J. Booij<sup>1</sup>

Received: 5 April 2023 / Accepted: 28 December 2023 / Published online: 9 January 2024  
© The Author(s) 2024

## Abstract

This study introduces and validates an artificial intelligence (AI)-based downscaling method for Standardized Precipitation Indices (SPI) in the northwest of Iran, utilizing PERSSIAN-CDR data and MODIS-derived drought-dependent variables. The correlation between SPI and two drought-dependent variables at a spatial resolution of 0.25° from 2000 to 2015 served as the basis for predicting SPI values at a finer spatial resolution of 0.05° for the period spanning 2016 to 2021. Shallow AI models (Support Vector Regression, Adaptive Neural Fuzzy Inference System, Feedforward Neural Network) and the Long Short-Term Memory (LSTM) deep learning method are employed for downscaling, followed by an ensemble post-processing technique for shallow AI models. Validation against rain gauge data indicates that all methods improve SPI simulation compared to PERSIANN-CDR products. The ensemble technique excels by 20% and 25% in the training and test phases, respectively, achieving the mean Determination Coefficient (DC) score of 0.67 in the validation phase. Results suggest that the deep learning LSTM method is less suitable for limited observed data compared to ensemble techniques. Additionally, the proposed methodology successfully detects approximately 80% of drought conditions. Notably, SPI-6 outperforms other temporal scales. This study advances the understanding of AI-driven downscaling for SPI, emphasizing the efficacy of ensemble approaches and providing valuable insights for regions with limited observational data.

## 1 Introduction

Drought, as a natural disaster, can impact various economic sectors, nature, and society. Currently, millions of people are facing challenges due to drought. Drought is defined as a period of precipitation that is below normal for months or even years. Furthermore, climate change may exacerbate the impacts of drought events with different characteristics. Although a drought typically begins with a precipitation shortage, scholars have identified different types of

drought, such as meteorological, hydrological, agricultural, socio-economic, and ecological drought (Crausbay et al. 2017), which can be interconnected. For example, groundwater is affected when meteorological drought persists for an extended period. This drought then leads to hydrological drought, resulting in a scarcity of soil moisture, also known as agricultural drought. Subsequently, socio-economic drought occurs when there is an imbalance between water supply and demand. When ecosystems are impacted by drought, it is referred to as ecological drought. Various drought types have distinct characteristics and quantification systems. To formulate effective policies, it is crucial to determine how drought has evolved and understand its potential impacts (Prodhan et al. 2021). However, many environmental studies have underestimated the significance of drought due to the lack of sufficient spatio-temporal field measurements of drought intensity at both regional and global scales (Shen et al. 2019). Spatial patterns of drought can be determined by applying interpolation methods to in-situ data. Nevertheless, accurate drought patterns cannot be captured by interpolation for a sparse in-situ network. Recent developments in space-based satellite techniques show promise, mainly due to the data measurement sensitivity throughout

✉ Nazanin Behfar  
n.behfar@utwente.nl

<sup>1</sup> Department of Water Engineering and Management, Faculty of Engineering Technology, University of Twente, 7500 AE Enschede, The Netherlands  
<sup>2</sup> Center of Excellence in Hydroinformatics and Faculty of Civil Engineering, University of Tabriz, 29 Bahman Ave., Tabriz, Iran  
<sup>3</sup> World Peace University, Sht. Kemal Ali Ömer Sok, Lefkosa, Turkey  
<sup>4</sup> College of Engineering, IT and Environment, Charles Darwin University, Darwin, Australia

the electromagnetic spectrum for subsurface soil moisture and drought conditions.

In recent times, there has been significant attention given to the use of satellite data for drought monitoring. Most of these efforts focus on the relationships between reflectance in the thermal and visible domains and drought events (Rhee et al. 2010; Small et al. 2018; Wang and Qu 2007). The fundamental concept underlying these efforts is that the optical and thermal domains can measure specific environmental variables, such as total water, surface temperature, vegetation fraction, and soil color, which are significantly impacted by drought events (Mohseni and Mokhtarzade 2020; Shen et al. 2019). Previous studies have proposed and examined various indicators for monitoring drought states across different climatic regions, including the standardized precipitation index (SPI), deciles index, Palmer drought severity index, and precipitation anomaly index. Among these, SPI stands out as the most widely used indicator for drought tracking and analysis. This is attributed to several factors: (1) its simple computation, (2) its utilization of available precipitation data, (3) its flexibility to compute for any required time scale, (4) its ability to assess both short-term and long-term consequences for water resources, particularly soil moisture, and (5) the main benefit of SPI lies in its applicability to study all types of drought. This is crucial considering the significance of precipitation in all these types of drought (Azimi and Azhdary Moghaddam 2020).

Although satellite-based data can be used to compute drought indices, precipitation variation usually occurs at smaller spatial scales compared to the size of the satellite pixel. Therefore, downscaling of satellite-based precipitation data will be an essential step for hydro-climatic modeling. Among different satellite products, precipitation estimation from satellite information using artificial neural networks—climate data record (PERSIANN-CDR) offers near real-time precipitation data employing microwave and infrared satellite records from global geosynchronous satellites. PERSIANN-CDR provides long-term information (since 1983) at a spatial scale of  $0.25^\circ$  in the  $60^\circ\text{S}$ – $60^\circ\text{N}$  latitude band (Ullah et al. 2019). PERSIANN-CDR is more credible compared to other spatial precipitation data because various data sources are utilized to create PERSIANN-CDR products (Ashouri et al. 2015). However, satellite precipitation estimates suffer from important drawbacks due to their coarse spatial scales (Sun et al. 2018), making them unsuitable for regional applications. Therefore, a reliable downscaling method is necessary to increase the spatial resolution of satellite precipitation estimates.

Methods of downscaling can be categorized as dynamical and statistical (Zhao et al. 2020). Dynamical downscaling methods provide fine-scale climatic data by modeling the physical processes based on a numerical weather model (NWM) or regional climate model (RCM), but they usually

require a high number of calculations. Dynamic methods are more suitable for downscaling meteorological variables present in RCMs (such as precipitation) rather than indices. The spatial resolution of RCMs varies from 20 to 60 km. Statistical downscaling methods simulate the statistical relation between fine-scale and coarse-scale covariates. In this way, auxiliary data can be employed to improve the downscaling performance (Immerzeel et al. 2009). Statistical downscaling methods are classified into weather typing and analogs, regression approaches, and weather generators. Due to their simplicity and low computational demand, regression methods are a good alternative to dynamical downscaling. Artificial intelligence (AI) methods, as a type of statistical downscaling method, are highly popular due to their high potential in modeling nonlinear and complex phenomena. Although various studies have been conducted to downscale precipitation (Goly et al. 2014; Sachindra et al. 2018; Tripathi et al. 2006), some of these studies have employed AI models to monitor and model drought. For instance, Neeti et al. (2021) developed an integrated framework employing a random forest algorithm to produce precipitation data at fine spatial scales and used the disaggregated results for monitoring meteorological drought. The results indicated that the developed framework is viable and generally adequate for application in other areas for drought analysis. Rhee and Im (2017) proposed a fine-scale drought modeling framework for ungauged regions. The standardized precipitation evapotranspiration index (SPEI) and SPI were estimated at a  $0.05^\circ$  spatial resolution using AI methods, as well as spatial interpolation by Kriging to satellite products. Park et al. (2017) developed a high-resolution soil moisture drought index (HSMDI) based on the random forest algorithm for monitoring hydrological, agricultural, and meteorological droughts. The obtained results showed a high correlation with in-situ data. Anagnostopoulou (2017) employed a statistical downscaling method using an artificial neural network (ANN) to simulate the SPI. RCM precipitation was also utilized to predict the SPI. The outcomes revealed that RCMs accurately captured the spatial extent, duration, and intensity of drought for SPI-12, while the ANN performed better for estimating SPI-3 compared to long-term indices. In another study, Anagnostopoulou et al. (2013) developed a statistical downscaling method using an ANN for simulating the SPI. While their methodology led to overestimated results for mean SPI, the regeneration of SPI-6 and SPI-3 for spring and winter seasons displayed reliable outcomes.

More recently, deep learning-based AI methods have been applied for drought modeling. For example, Lee et al. (2019) used a deep feedforward neural network (DFNN) model to simulate soil moisture via satellite data for drought investigation, showing a high correlation with in-situ measurements. Zhang et al. (2017) found that a DFNN model trained by satellite data is capable of modeling complicated patterns

of soil moisture. Agana and Homaifar (2017) designed a deep belief network (DBN) to model drought using only historical data of the standardized streamflow index (SSI) as predictors. Shen et al. (2019) proposed a drought monitoring method employing a DFNN model, demonstrating a high ability to simulate agricultural and meteorological droughts. Prodhan et al. (2021) monitored agricultural drought employing a DFNN where soil and vegetation parameters, as well as precipitation, were considered as predictors. They showed that the DFNN method outperformed distributed random forest (DRF) and gradient boosting machine (GBM) models. Furthermore, several studies have used deep learning to downscale precipitation. For instance, Kumar et al. (2021) employed a methodology based on a convolutional neural network (CNN) to downscale precipitation over India, showing high accuracy compared to observations. In another study by Sun and Lan (2021), precipitation downscaling was performed using a CNN model over China, revealing that the CNN model had high performance and good skills in reproducing frequency distributions of daily precipitation. Wang et al. (2021) suggested a super-resolution deep residual network (SRDRN) for precipitation downscaling. The outcomes demonstrated that the SRDRN method not only captured spatial and temporal patterns remarkably well but also reproduced precipitation extremes in different locations and times at the local scale. However, it is essential to note that deep learning models have a high number of trainable parameters, and in situations where data is scarce, a deep learning model may not be a suitable choice (Sharghi et al. 2022).

Despite the fact that data-driven methods (e.g., ANNs, etc.) may provide relatively reliable results, it is clear that for a given problem, different methods may yield different results. Thus, by integrating different methods through an ensemble technique, different patterns of phenomena can be captured more accurately. One specific method is not necessarily the best for all time periods and conditions. Such model ensemble methods have been used in various contexts (e.g., see Nourani et al. 2022; Shamseldin et al. 1997; Sharghi et al. 2019), but to the best knowledge of the authors, not in the field of downscaling and drought modeling. Additionally, there is no study that has applied deep learning-based downscaled precipitation products for spatio-temporal drought monitoring.

In this paper, to fill the aforementioned gaps, an attempt was made to reproduce SPI (as a simple and commonly used index for drought monitoring) at a fine spatial resolution based on PERSIANN-CDR data. The suggested framework not only downscales precipitation but also predicts the SPI. Moreover, the vegetation fraction index and soil temperature were utilized as drought-dependent variables to downscale PERSIANN-CDR-based SPI employing shallow AI methods (adaptive neural fuzzy inference system (ANFIS), support vector regression (SVR), feedforward neural network

(FFNN)), and the deep learning-based method of long short-term memory (LSTM). Finally, an AI-based modeling ensemble technique was used as a post-processing method.

## 2 Used materials and methods

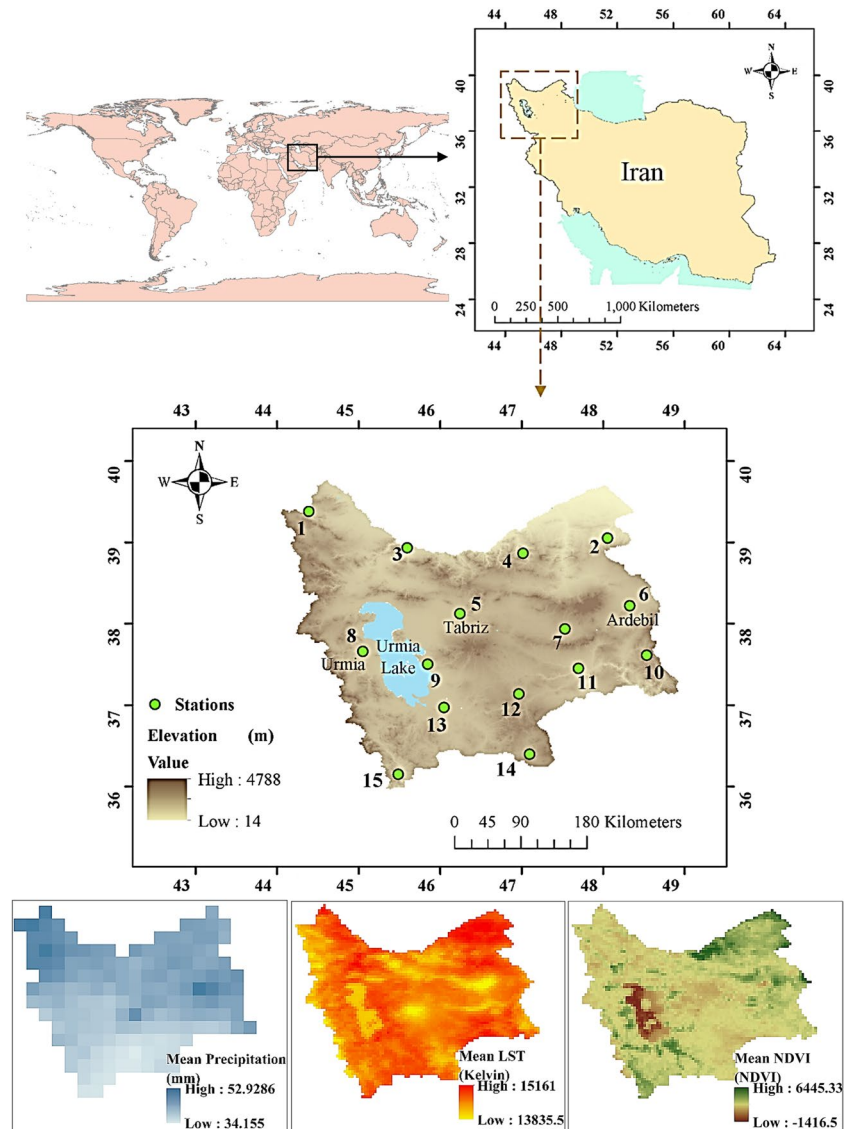
### 2.1 Study area and employed data

#### 2.1.1 Study area and in-situ data

The proposed methodology has been applied to the north-western region of Iran. Figure 1 presents the spatial extent and geographic position of the study area, which includes two lakes and seventeen rivers. The northern, snow-capped district, incorporating Lake Urmia, experiences abundant precipitation, characterized by profound valleys and fertile lowlands. The climate in the western areas is mountainous due to numerous elevations. The Zagros Mountains, extending from northwest to southeast, hinder the full flow of moisture-laden currents from the Atlantic Ocean and the Mediterranean Sea. Additionally, cold Siberian and Central Asian air currents entering the western part from the north and northeast intensify the coldness and humidity of the climate, leading to severe cold and heavy snowfall. Another influential factor in the region's climate is the presence of Lake Urmia, which not only provides humidity but also moderates temperatures. The maximum temperature reaches 34 °C in July, while the minimum temperature drops to −16 °C in January. Generally, the mean annual precipitation in the western area is 300–400 mm. The volcanic peak of Sahand Mountain (3707 m) is the highest point in the central area. It has a semi-arid and cold climate influenced by the Mediterranean Continental air condition. The low-lying regions are also affected by gentle breezes from the Caspian Sea. The central area is a cold and mountainous region, considered a semi-arid area with a mean annual precipitation of 250 to 300 mm. The eastern area, situated in a mountainous region, experiences a cold, semi-arid climate. Large parts of the eastern area are green and forested, with bitterly cold winters and temperatures dropping to −25 °C. The special geographical and topographic features of the eastern area, including mountain ranges exceeding 4000 m in altitude and wide plains, contribute to better precipitation conditions compared to other parts of the country, ranging between 250 and 600 mm.

Northwestern Iran encompasses various meteorological sites with several in-situ stations in operation. Among the various measurements taken by these stations (precipitation, evaporation, and air temperature), this paper utilized a monthly precipitation dataset from 15 stations (see Fig. 1) to evaluate the computed high-resolution SPI indices via the proposed AI-based downscaling method.

**Fig. 1** Study area, selected meteorological sites and mean monthly precipitation, LST and NDVI maps (2015–2021)



### 2.1.2 Satellite data

In this study, monthly PERSIANN-CDR data obtained from <https://chrsdata.eng.uci.edu/> were utilized. Derived from satellite infrared measurements, PERSIANN-CDR is a collaborative effort between the National Oceanic and Atmospheric Administration (NOAA) and the Center for Hydrometeorology and Satellite (CHRS). It is employed to generate spatially distributed daily precipitation data from 1983 to the present. Unlike previous PERSIANN products, such as real-time and adjusted versions, PERSIANN-CDR utilizes GridSat-B1 information rather than passive microwave and Climate Prediction Center datasets, thereby using a different infrared dataset as input. Notably, PERSIANN-CDR provides long-term continuous precipitation data since 1983, rendering it more suitable for drought analysis that requires long-term datasets (Ashouri

et al. 2015). Data retrieval for PERSIANN-CDR utilizes ANNs and provides a spatial scale of  $0.25^\circ$  with quasi-global coverage ( $60^\circ\text{N}$ – $60^\circ\text{S}$ ). Additionally, the 3-hourly precipitation rates are accumulated to determine monthly precipitation (Mohseni et al. 2021).

In this study, two satellite-based datasets were used as auxiliary data alongside the precipitation data. Day-time Land Surface Temperature (LST) measurements ( $0.05^\circ \times 0.05^\circ$ ) served as the satellite-based soil temperature parameter. Furthermore, the Normalized Difference Vegetation Index (NDVI) obtained from MOD13C2 measurements was used as the satellite-based vegetation factor. The satellite products utilized in this paper include Terra Monthly Vegetation Indices (MOD13C2) and Terra Monthly LST & Emissivity (MOD11C3) from MODIS measurements at a resolution of  $0.05^\circ$ . The MOD11C3 data provides monthly mean LST & emissivity values.

## 2.2 Standard precipitation index

The standard precipitation index (SPI) is an indicator utilized to characterize environmental drought across various timescales, including 3, 6, and 12 months. Vegetation conditions and soil thermal inertia serve as proxies for hydrological and agricultural droughts influenced by vegetation water and soil moisture. In short periods, such as one month, a decrease in precipitation may not impact soil thermal inertia or vegetation health. However, prolonged periods of reduced precipitation can affect surface water storage and vegetation health over a season. SPI, calculated from precipitation data, exhibits a strong correlation with vegetation and temperature conditions. To generate high-resolution environmental drought indices, SPI can be downscaled from coarse-scale satellite estimations of precipitation using MODIS products. In this study, we explored the correlation between two drought-dependent parameters, LST and NDVI, and SPI for downscaling environmental drought at a fine spatial scale.

The SPI is computed from precipitation data at different timescales, including 3, 6, and 12 months, using the formula:

$$SPI = \frac{X - \bar{X}}{\sigma} \quad (1)$$

where  $x$  is the precipitation value,  $\bar{X}$  is the average precipitation over a specific period, and  $\sigma$  is the standard deviation of precipitation. In this study, SPI was calculated in R Studio using the SPEI package for the period 2000 to 2021. It is important to note that different SPI indices are considered due to their application in exploring distinct objectives. For instance, in agricultural drought assessment, each index is valuable for investigating the impact on different crops. Calculating SPI at smaller temporal scales (e.g., 3 months) can serve as an indicator for immediate effects such as decreased flow, snowpack, and soil moisture in smaller streams. Conversely, at medium scales (e.g., 6 to 12 months), it can indicate decreased stream flow and reservoir storage. Therefore, to obtain a comprehensive understanding of the potential effects of a drought, SPI should be computed at different time scales. The correlation between the time scale and the effect of a drought is contingent on human interference and natural conditions. As discussed in previous studies, drought indices, including SPI, are often correlated with soil temperature and vegetation (Zhu et al. 2017). Various drought classifications exist in the literature based on different SPI thresholds. Table 1 illustrates the classification of meteorological drought by McKee et al. (1993).

It is important to mention that auxiliary data, specifically NDVI and LST, were also transformed to match the same specific periods used for calculating the SPIs. The newly computed variables are denoted as NDVI-3, NDVI-6, NDVI-12, LST-3, LST-6, and LST-12.

**Table 1** Drought classification based on SPI scores (McKee et al. 1993)

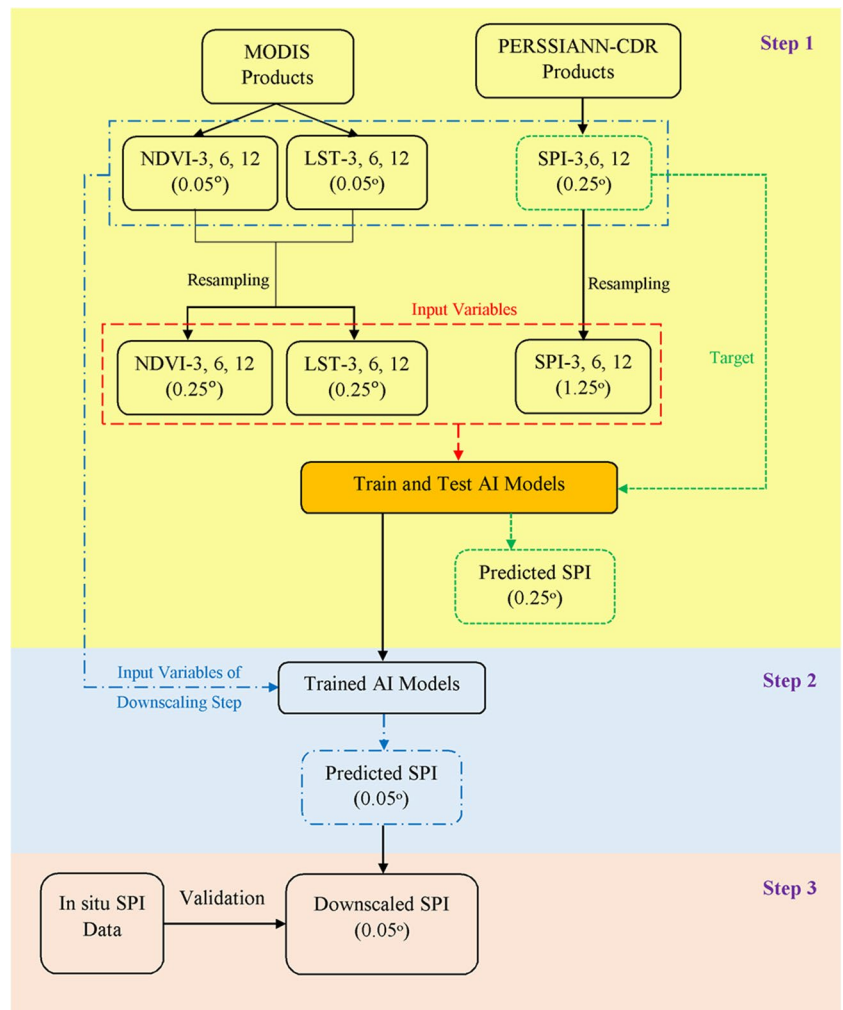
SPI value	Drought class
$SPI > 2.0$	Extremely wet
$1.5 \leq SPI < 2.0$	Very wet
$1.0 \leq SPI < 1.5$	Moderate wet
$-1.0 \leq SPI < 1.0$	Normal
$-1.5 \leq SPI < -1.0$	Moderate drought
$-2.0 \leq SPI < -1.5$	Severe drought
$SPI < -2.0$	Extreme drought

## 2.3 Downscaling methodology

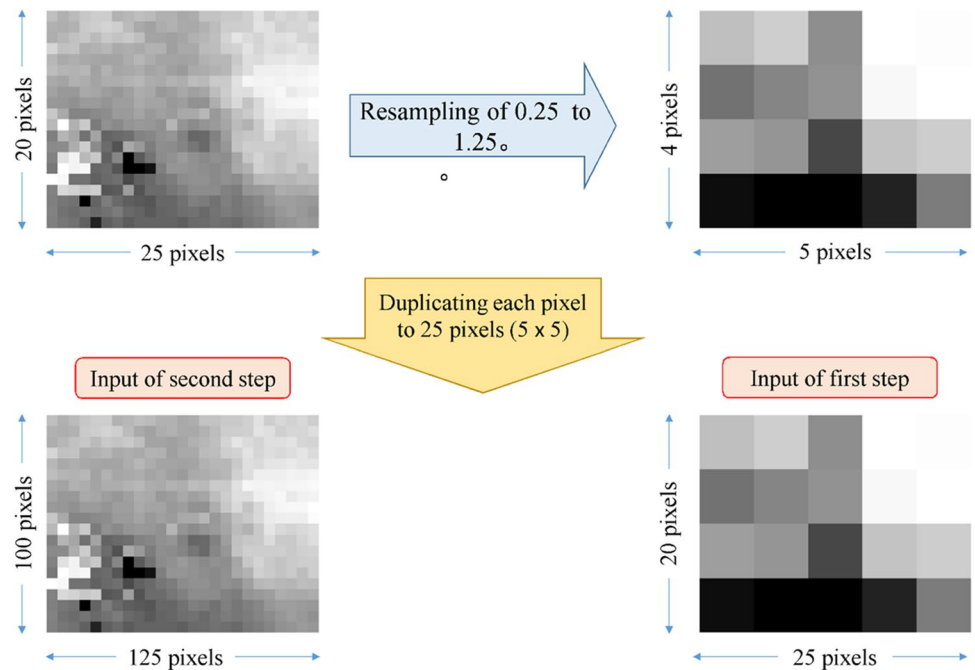
In recent decades, various statistical downscaling methods have been employed to downscale satellite-based hydro-climatic data. AI methods have demonstrated acceptable performance across areas with diverse topographic and climatic conditions. Figure 2 outlines the main steps of the proposed downscaling framework in this study. Shallow AI models, including FFNN, a commonly used AI method; ANFIS, a hybrid AI method capable of handling uncertainties through fuzzy concepts; SVR, an AI model minimizing operational risks; and the deep learning model of LSTM, increasingly popular but needing precise comparison with shallow AI models, are employed. An ensemble technique is also incorporated to downscale the SPI at different temporal scales. The methodology operates under the assumption that the statistical relationship between NDVI and LST variables as predictors and SPI at a lower resolution remains valid at a higher resolution. Previous studies, such as Gidey et al. (2018), have considered and confirmed the validity of this assumption. To establish the connection between the  $0.05^\circ$  predictive variables and the  $0.25^\circ$  SPI data, NDVI and LST parameters were initially resampled to a resolution of  $0.25^\circ$  using the bilinear method. Simultaneously, the  $0.25^\circ$  SPI data were resampled, using the bilinear technique, to  $1.25^\circ$  for use as input in the modeling. This involved dividing each pixel of  $1.25^\circ$  into a  $5 \times 5$  grid to match the  $0.25^\circ$  resolution.

The primary objective of this study was to compute SPI data at a resolution of  $0.05^\circ$ , utilizing SPI data both as input and target. Given the original resolution of  $0.25^\circ$ , the proposed approach involved training the model at this resolution to facilitate downscaling to  $0.05^\circ$ . To achieve this, initially the model was trained to downscale from  $1.25^\circ$  to  $0.25^\circ$ , leveraging datasets available at both resolutions. Once the model was trained successfully at the  $0.25^\circ$  resolution, it was applied to downscale from  $0.25^\circ$  to  $0.05^\circ$ , as data at the finer resolution was not directly available for training purposes. Considering that the downscaling factor is 5 times ( $0.25^\circ$  to  $0.05^\circ$ ), this process was performed to accurately compute SPI data at the desired  $0.05^\circ$  resolution (refer to Fig. 3). This

**Fig. 2** Overall diagram of the proposed framework for downscaling SPI



**Fig. 3** Resampling of SPI data from 0.25° to 1.25° to be used as input of models to improve modeling performance



multi-step training approach allowed it to effectively utilize available data and bridge the resolution gap, enabling the computation of SPI data at the targeted finer resolution. Finally, to assess the modeling performance, the simulated results were compared with the recorded gauge-based data. For further clarification, information on the employed satellite data is tabulated in Table 2.

The sub-steps of the proposed methodology for downscaling SPI are as follows:

- i. LST and NDVI products from MODIS at  $0.05^\circ$  were resampled to a  $0.25^\circ$  scale using the bilinear interpolation method to serve as predictors for the models.
- ii. PERSIANN-CDR products at  $0.25^\circ$  were resampled to  $1.25^\circ$  using the bilinear method, and the precipitation data were used to calculate SPI as a predictor for the models.
- iii. The models were trained using monthly data from 2000 to 2015, while the data from 2016 to 2021 were used for validation purposes.
- iv. The trained models were employed to estimate SPI values at  $0.05^\circ$  using LST and NDVI datasets at a  $0.05^\circ$  resolution for the years 2020–2021.
- v. Validation of SPI downscaled to a resolution of  $0.05^\circ$  was carried out employing in-situ data.

In some similar studies, SPI was not included as an input, and the residuals were interpolated at the end to reduce errors. However, in this study, by considering SPI as an input, there is no need for the final interpolation of residuals because the upscaled SPI has been used in the model training from the beginning to enhance the modeling (Mohseni et al. 2021; Mohseni and Mokhtarzade 2020; Sharifi et al. 2019).

### 2.3.1 Feed-forward neural network

The feed-forward neural network (FFNN) method, calibrated by the backpropagation (BP) algorithm, is the most commonly used model in various engineering studies. The first form of ANNs is the FFNN, where the phrase “feed-forward” signifies the forward flow of data during the network’s training.

- a) Various FFNN architectures, ranging from one middle layer (with 1 to 10 neurons), and different epoch numbers were created and tested.
- b) The model was calibrated using the BP algorithm, applying gradient descent optimization.
- c) The model demonstrating the best performance, i.e., the model with the maximum determination coefficient, was selected.

It should be mentioned that there is no specific method to design an ANN; therefore, each model should be structured individually.

### 2.3.2 Adaptive neural fuzzy inference system

The term “neuro-fuzzy” modeling refers to the process of utilizing different learning algorithms to apply fuzzy simulation to neural networks or fuzzy inference systems. Fuzzy systems comprise three main components: a fuzzifier, a fuzzy database, and a defuzzifier. It should be noted that the two main parts of a fuzzy database are the fuzzy rule base and the inference engine. A fuzzy rule base consists of rules that relate to fuzzy propositions (Jang et al. 1997). Therefore, fuzzy inferences are employed in operational analysis. The Sugeno FIS used by Jang (1993) was applied in this study.

### 2.3.3 Support Vector Regression

Cortes and Vapnik (1995) introduced the support vector machine (SVM) approach, offering a satisfactory solution for problems such as pattern recognition, regression, prediction, and classification. Statistical learning theory and structural risk minimization are two useful features of SVM, distinguishing it from FFNN. SVR, a type of SVM, entails fitting a linear function to datasets and subsequently applying a nonlinear kernel to the previous results to simulate the target data in a nonlinear manner (Raghavendra and Deka 2014).

### 2.3.4 Ensemble unit

A model ensemble method is an approach that combines different techniques to enhance the final accuracy. Various

**Table 2** Data information employed in this study

Dataset	Sources	Temporal resolution	Spatial resolution (original)	Resampled resolution as input	Computed output resolution
Precipitation	PERSIANN-CDR of CHRS	Monthly	$0.25^\circ$	$1.25^\circ$	$0.05^\circ$
LST	MOD11C3 of MODIS	Monthly	$0.05^\circ$	$0.25^\circ$	-
NDVI	MOD13C2 of MODIS	Monthly	$0.05^\circ$	$0.25^\circ$	-

studies in diverse contexts have recommended ensembling the results of various models as an efficient post-processing method to improve estimation performance (Kazienko et al. 2013). The use of ensemble methods in this study is justified by their capacity to capitalize on individual models' strengths while alleviating their weaknesses. The rationale includes: 1) Diversity of Models: Ensembles merge predictions from diverse models, capturing a broader range of patterns. 2) Compensating for Model Limitations: Ensembles offset individual model limitations, excelling in various regions or conditions. 3) Reduction of Overfitting: Nonlinear neural ensembles reduce overfitting, enhancing generalization to unseen data. 4) Improved Robustness: Ensembles minimize outliers' impact, valuable in complex environmental modeling. 5) Performance Boost: Combining model outputs often yields superior performance. 6) Model Consensus: Ensembles generate predictions based on consensus, enhancing reliability. In summary, ensembles were strategically chosen to enhance downscaling methodology, providing a more comprehensive and accurate representation of underlying data relationships, addressing variability and uncertainties in environmental modeling.

In this paper, a nonlinear technique was employed to integrate the outcomes of the individual models used to enhance the final modeling accuracy. A new FFNN is created as a nonlinear ensemble method. The results of single methods (SVR, ANFIS, and FFNN) are considered as inputs for another new FFNN model to create the nonlinear neural ensemble model (refer to Fig. 4).

It should be noted that the employed ensemble technique is a shallow AI ensemble, which combines shallow single AI models to utilize each model's capability, enhance shallow modeling accuracy, and reduce uncertainty. Shallow AI models, having fewer layers, may face difficulty in analyzing more complicated tasks. However, deep learning models with more layers can extract features and patterns from datasets more accurately and easily. But deep learning methods require more data for training, and in the case of limited data availability, an ensemble post-processing technique may be a better choice.

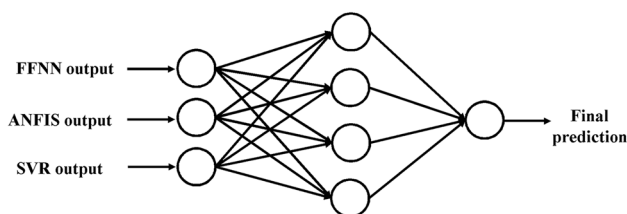


Fig. 4 Diagram of suggested nonlinear ensemble method

### 2.3.5 Long short-term memory (LSTM) network

Sequences with long-short-term temporal information cannot be effectively recognized by conventional neural networks. To address this issue, a special type of recurrent neural network (RNN) has been designed, employing a loop framework to learn the temporal dependence between different variables. However, a typical RNN struggles to capture long-term dependencies effectively, as it tends to lose current and dependent information. A specific RNN structure, known as the LSTM network (Hochreiter and Schmidhuber 1997; Li et al. 2022), has been developed to overcome this limitation. In the LSTM network, three “gate” structures are designed to create a chain architecture of memory units. Figure 5 illustrates the diagram of a long short-term memory (LSTM) block, featuring, for example, 5 hidden units and 2 input dimensions.

Through a ‘gate’, information is either discarded or retained within the memory unit to safeguard and control it. The LSTM network effectively addresses both the vanishing and exploding gradient problems through its memory unit, overcoming the limitations of conventional RNNs. In this study, the LSTM network comprises a sequence input layer, LSTM layer, fully connected layer, and regression layer.

### 2.4 Modeling performance

The root mean square error (RMSE) and Nash–Sutcliffe efficiency (or determination coefficient, DC) were employed to evaluate the modeling performance (Nourani and Behfar 2021):

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (I_{(t)} - I_{com(t)})^2}{N}} \quad (2)$$

$$DC = 1 - \frac{\sum_{t=1}^N (I_{(t)} - I_{com(t)})^2}{\sum_{t=1}^N (I_{(t)} - \bar{I})^2} \quad (3)$$

where  $N$  stands for the number of data points,  $I_{(t)}$  is the observed variable,  $\bar{I}$  is the average of the observed variable, and  $I_{com(t)}$  is the modeled variable. DC ranges between  $-\infty$  and 1, with an optimal value of 1, while RMSE ranges between 0 and  $\infty$ , with an optimal value of 0. It should be noted that DC is different from the correlation coefficient (CC). DC indicates the agreement between modeled and observed variables, while CC indicates the linear correlation between them.

Although DC, RMSE, and CC are the most commonly used criteria for the verification of satellite data, as



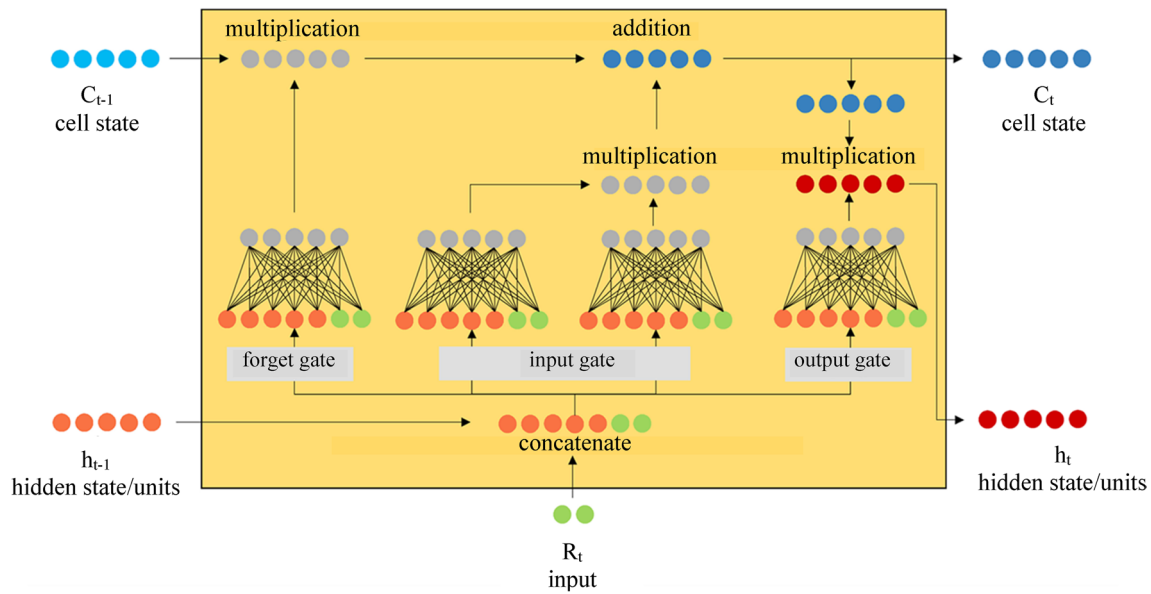


Fig. 5 Structure of LSTM unit with 5 hidden units and 2 input dimension

reported by Baez-Villanueva et al. (2018), RMSE alone is not sufficient for the verification of satellite-based precipitation data. Thus, for a more precise verification, the false alarm ratio (FAR) and probability of detection (POD) could also be utilized for verification of the precipitation data:

$$POD = \frac{H}{H + M} \tag{4}$$

$$FAR = \frac{F}{H + F} \tag{5}$$

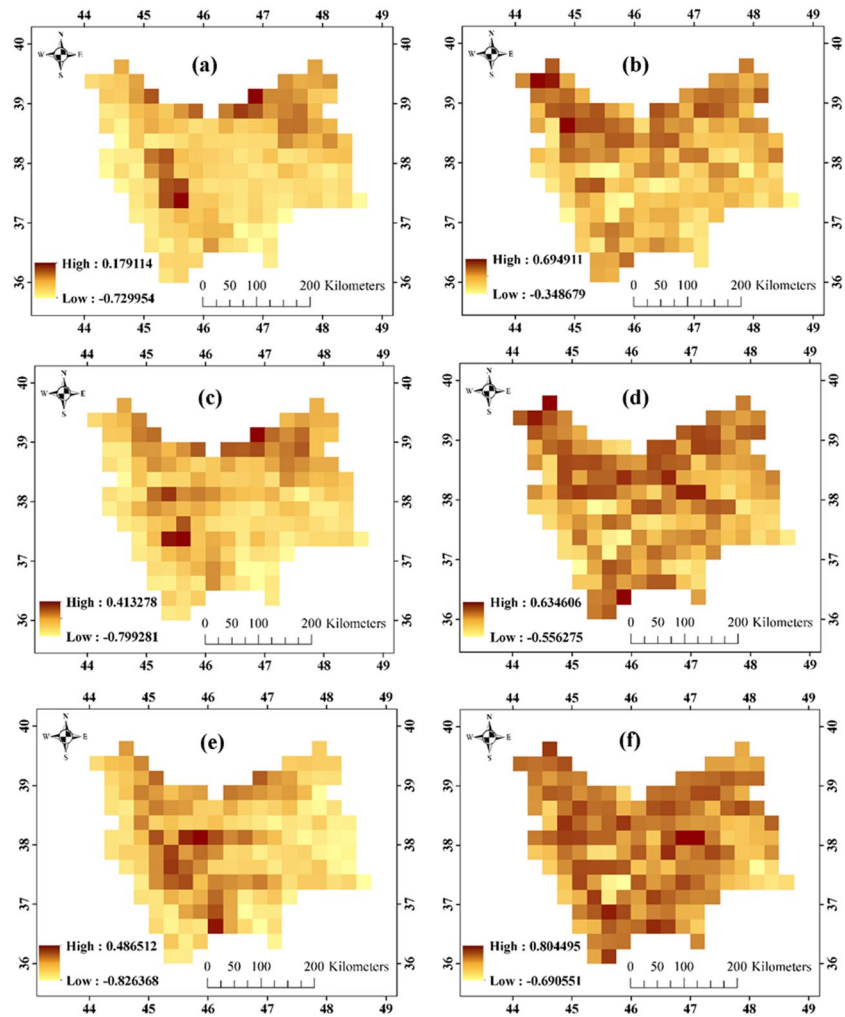
where  $H$  is the number of satellite-based precipitation measurements that correctly detected the drought event ( $SPI < -1$ ),  $M$  represents the number of drought events reported by the ground station but not detected by the satellite-based precipitation estimates, and  $F$  refers to the number of drought conditions identified by the satellite-based precipitation estimates but not reported at the ground station. Overall, the FAR and POD scores range from 0 to 1, with optimal values of 1 and 0 for POD and FAR, respectively.

### 3 Results and discussion

In this study, initially, NDVI and LST were upscaled to a resolution of  $0.25^\circ$  employing the bilinear resampling method. The framework operates under the assumption that the methods (FFNN, ANFIS, SVR, and LSTM) linking the SPI to NDVI and LST at a low spatial resolution ( $0.25^\circ \times 0.25^\circ$ ) are

equally effective at a high spatial resolution ( $0.05^\circ \times 0.05^\circ$ ; Zhang et al. 2018). The effectiveness of these methods depends on the relation between the noted datasets (Zhang and Jia 2013; Mohseni et al. 2021). According to the thermal inertia theory, as vegetation intensity decreases, soil temperature increases, and vice versa. Therefore, the correlation between vegetation fraction and soil temperature is negative. This negative relation is more prevalent under dry soil conditions. In this context, SPI has a positive relation with NDVI and a negative relation with LST. As soil temperature decreases, the drought indices increase and vice versa. Some studies have reported that LST is more related to SPI compared to NDVI (e.g., Mohseni et al. 2021), while others (e.g., Zhang and Jia 2013) reported the reverse case. This discrepancy can be related to the climate and vegetation intensity of a region. In the study area, where vegetation is not very dense, the correlation of SPI with NDVI was less apparent than that with LST. Conversely, evaporation is high in this region, which is characterized by land temperature. Thus, the correlation of SPI with LST was higher than with NDVI. The degree of correlation between a drought index and vegetation/soil temperature variables not only depends on climatic conditions but also is affected by the type of drought index. For instance, SPEI is strongly dependent on evapotranspiration, but employing such indices requires more datasets. A correlation analysis was conducted among LST, NDVI, and SPI at different scales (Fig. 6). In Fig. 6, the left column shows the CC pattern of LST and SPI at different time scales, while the right column depicts the CC pattern of NDVI and SPI over the study area. The spatial distribution of CC shows almost similar patterns at different time scales

**Fig. 6** Correlation between SPI-3 and: (a) LST, (b) NDVI, SPI-6 and: (c) LST, (d) NDVI, SPI-12 and: (e) LST, (f) NDVI



influenced by the climate regimes of different areas. According to Fig. 6, in northern regions, the relationship of SPI with LST is weaker than in southern regions, but the relationship of SPI with NDVI is stronger than in other regions. Conversely, in the central and southern regions, SPI has a stronger relationship with LST and a less strong relationship with NDVI compared to the northern regions.

The linking models (FFNN, ANFIS, SVR, and LSTM) between low-resolution PERSIANN-CDR data-based SPIs (i.e., SPI-3, SPI-6, and SPI-12) and NDVI and LST were created using a training dataset. For the FFNN model, different networks with varying numbers of neurons and epochs were examined to select the optimum network structure for down-scaling. In the case of ANFIS, different membership functions (MFs) were employed, with the triangular MF yielding better performance in the simulation of SPI. The variation of the calibration iterations (epochs) was also explored to determine the best ANFIS method. The SVR method was implemented using the radial basis function (RBF) kernel. The number of tuning factors for the RBF kernel is smaller than for both polynomial and sigmoid kernels. Additionally,

based on smoothness assumptions, the RBF kernel could lead to better performance compared to other kernel functions (Noori et al. 2011). The parameters of the RBF kernel in SVR models were tuned to obtain optimum results (see Table 3). As mentioned earlier, the upscaled SPI data, as well as LST and NDVI, were imposed into the input layers of models to enhance modeling accuracy. The outcomes of the first stage are summarized in Table 3. In addition to the shallow learning AI models, deep learning-based modeling was also carried out via the LSTM model. To design the structure of the LSTM model, a trial-and-error procedure was applied to obtain the optimum network.

Next, outputs of single shallow AI models (FFNN, ANFIS, and SVR) produced in the previous stage were integrated using the ensemble post-processing method. In a nonlinear ensemble method similar to the single FFNN, a network was trained utilizing the scaled conjugate gradient method of the BP algorithm and using the tan-sig activation function for hidden and output layers. Also, a trial and error procedure was employed to obtain the optimal structure. The results of the ensemble post-processing technique are

**Table 3** Results of the first step for single shallow, deep learning, and ensemble methods for training and test phases

Scale	Model	Structure <sup>1,2</sup>	DC		RMSE (Non dimensional)	
			Train	Test	Train	Test
SPI-3	FFNN	3	0.6473	0.6452	0.5099	0.5505
	ANFIS	trimf-2	0.6390	0.6381	0.5159	0.5560
	SVR	0.33–0.1–1	0.6541	0.6503	0.5050	0.5465
	Ensemble	9	0.7728	0.7700	0.4405	0.4117
	LSTM	5	0.8740	0.6618	0.3902	0.5295
SPI-6	FFNN	5	0.6576	0.6447	0.4601	0.4786
	ANFIS	trimf-3	0.6932	0.6915	0.4302	0.4459
	SVR	0.01–0.01–1	0.6852	0.6834	0.4370	0.4517
	Ensemble	3	0.8283	0.7965	0.3633	0.3539
	LSTM	5	0.9113	0.6932	0.3168	0.4299
SPI-12	FFNN	3	0.6124	0.5098	0.5446	0.7229
	ANFIS	trimf-2	0.6563	0.5410	0.5129	0.6871
	SVR	0.33–0.1–7	0.6479	0.5267	0.5191	0.7139
	Ensemble	5	0.7179	0.6667	0.4646	0.4068
	LSTM	5	0.8844	0.5467	0.3738	0.6806

<sup>1</sup> Inputs of FFNN, ANFIS, SVR, and LSTM models consist of LST, NDVI, and upscaled SPI

<sup>2</sup> Inputs of Ensemble model consist of outputs of ANFIS, FFNN, and SVR models

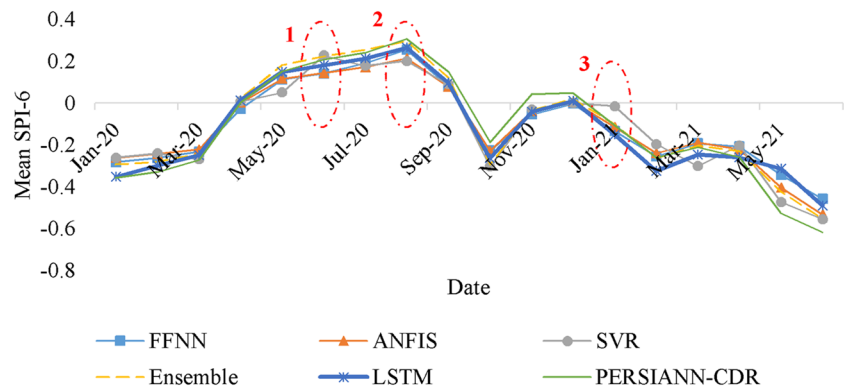
presented in Table 3 as well. It should be noted that a, b, and c in the SVR method stand for the constant parameter (a), approximation accuracy (b), and kernel parameter (c). In the ANFIS structure, MF-a refers to employed MFs and the number of MFs (a). Furthermore, a in the structure of an FFNN and ensemble refers to the number of intermediate neurons (a). Also, a in the LSTM model represents the number of hidden units. Coding was performed in MATLAB software to execute AI methods.

The presented results in Table 3 could be interpreted to follow the aim of the study, which is to explore the capability of deep learning in comparison to shallow AI-based models. Additionally, it aims to evaluate the impact of a nonlinear ensemble post-processing technique in improving the modeling results and eventually compare deep learning and shallow AI-based ensemble methods. As evident from Table 3, the outputs of the individual shallow AI methods are close to each other. However, at different stages and time scales, one of the models slightly outperformed the others. In general, ANFIS, due to the use of fuzzy theory, performed a little better than others in most cases. According to Table 3, the ensemble method could enhance the results of single shallow AI models for SPI-3, SPI-6, and SPI-12 by 18%, 19%, and 9%, respectively, for the training phase, and 18%, 15%, and 23%, respectively, for the test phase. It is evident that the improvement of DC in the training phase is not significant for SPI-12, but the performance improvement is considerable in the test phase for all temporal scales. The modeling improvement of the test phase for SPI-12 is more than for others, considering that the overall results of

SPI-12 are less accurate than for others. Considering that each model has its own advantages and disadvantages, the ensemble technique, because it incorporates the benefits of all models, may simulate the task more precisely than the single methods. For instance, the mean SPI-6 obtained using PERSIANN-CDR products and the corresponding mean predicted SPI-6 by employed models over January 2020–June 2021 are depicted in Fig. 7, where it is clear that the ensemble excels the other methods. For instance, points 1, 2, and 3 in June 2020, August 2020, and January 2021, respectively, were considered. Considering the noted points, SVR, FFNN, and ANFIS outperformed other shallow AI models at points 1, 2, and 3, respectively. Overall, the ensemble technique showed the best performance among all employed methods. It should be noted that there are different methods for combining model outputs, but as discussed in previous studies (Shamseldin et al. 1997; Sharghi et al. 2019), the neural ensemble model generally outperforms others. Therefore, in this study, the neural ensemble method was employed to improve the modeling performance.

On the other hand, the deep learning model LSTM significantly excels over the shallow AI models in the training phase, while it tends to slightly outperform the shallow AI models in the test phase, and overtraining of LSTM is evident. This is due to the high number of trainable parameters of deep learning models. For instance, FFNN, ANFIS, SVR, and LSTM models for modeling SPI-3 have 16, 50, 3, and 186 trainable parameters. In other words, because the LSTM model has many degrees of freedom, it has high performance in the training phase, but the presence of errors

**Fig. 7** Time-series comparing mean SPI-6 derived from PERSIANN-CDR data and corresponding mean predicted SPI-6 by employed models from January 2020 to June 2021



in multiple calibrated parameters increases the error in the testing phase. Also, LSTM results are more deteriorated for SPI-12, which has less training data compared to other indices. At this point, it is worth mentioning that the number of calibration data is very important. Hence, if more data are available and used for training the models, the deep learning model may perform significantly better than other models. Thus, the ensemble post-processing method can be a good alternative to a deep learning method when limited data are available. On the other hand, it is clear from Table 3 that the performance of methods in estimating SPI-3 and especially SPI-6 is higher than in estimating SPI-12.

In the second step, fine-scale inputs were imposed into the trained single shallow AI models as well as the LSTM to downscale SPI. The results of this step are presented in Table 4. It is worth mentioning that in some previous downscaling studies (e.g., Sharifi et al. 2019), residual correction was employed to correct and improve the downscaling outcomes. However, considering that in this study, the upscaled SPI was employed as input data, there is no need for residual correction. In other words, in previous downscaling studies, only auxiliary parameters were utilized as input for modeling, which is why the results were not so good. Therefore, residual correction should be done to correct the bias of estimations. In fact, the studies that only used auxiliary parameters as input do not exactly address the downscaling in the training phase of the modeling. Instead, in the last step, the residuals of estimated results are corrected by the low-resolution target. In contrast, in the current modeling, when the low-resolution target is used as an input, the error is reduced by the BP algorithm and loss function via the training process, and there is no need to modify the residuals as a further modeling step. Comparing the employed methodology in this study and previous studies, it is clear that the utilized methodology in some previous studies consists of more steps (i.e., one interpolation by ANN, one interpolation by spline, and again, the sum of ANN estimations and resampled errors, e.g., see Sharifi et al. 2019) than the methodology proposed in this paper (i.e., only one interpolation by AI model).

Finally, an evaluation was conducted by comparing the computed SPI-3, SPI-6, and SPI-12 indices with the observed data from ground gauges and the predicted SPI indices by the proposed method to provide a more detailed and in-depth comparison of the employed models. The results of this evaluation and comparison are presented in Table 4. Considering the results presented in Table 4, it can be interpreted that the performance of modeling for both SPI-3 and SPI-6 indices is almost similar. Moreover, regarding the obtained DC, the suggested model shows better results in estimating SPI-6 compared to the SPI-3 values. POD values computed for SPI-6 are 0.76 and 0.8 for LSTM and ensemble methods, respectively, indicating that about 80% of the drought conditions could be detected by the proposed modeling framework. On the other hand, FAR scores of SPI-6 are 0.22 and 0.24 for ensemble and LSTM methods, respectively, indicating that 25% of the drought events captured by the methodology are not real. Also, Fig. 8 shows the downscaled results of ensemble and LSTM methods for SPI-6 for a wet month (September 2020) over the study area with a mean SPI of 0.32. It should be noted that the study area is a semi-arid area, and there are not many months with an SPI greater than 1. Furthermore, Fig. 8 depicts similar outcomes for a dry month (June 2021), in which the mean SPI is about -1.3. The computed SPI by the PERSIANN-CDR data is also displayed to compare the results of deep learning and ensemble methods. According to Fig. 8a, the PERSIANN-CDR's SPI data vary from -1.02 to 1.64 in September 2020 as a wet month. The mean SPI is 0.31 for this month. In June 2021, drought conditions can be observed since the majority of SPI data in this month are negative (Fig. 8d). Considering Fig. 8, it is obvious that the low-resolution SPI map derived from PERSIANN-CDR data and high-resolution maps obtained from the proposed methodology are almost in the same range in both dry and wet months, indicating the competency of the suggested framework in linking vegetation and thermal factors to the drought indices. However, the estimates of the ensemble technique were more accurate and reliable than those of the LSTM model.

**Table 4** Results of the downscaling phase of the modeling for individual shallow, deep learning and ensemble methods at different temporal scales of SPI

Data scale		SPI-3						SPI-6						SPI-12					
		DC	CC	RMSE	POD	FAR	DC	CC	RMSE	POD	FAR	DC	CC	RMSE	POD	FAR			
		0.25°	PERSIANN- CDR's SPI	0.4864	0.7244	0.6992	0.7302	0.2698	0.6537	0.8086	0.5502	0.7273	0.1724	0.6654	0.8175	0.4529	0.6757	0.2308	
	FFNN	0.3411	0.5674	0.7619	0.6032	0.2083	0.4512	0.6727	0.6927	0.6364	0.2364	0.2723	0.5594	0.6679	0.3784	0.3778			
	ANFIS	0.3342	0.5622	0.7660	0.6032	0.1915	0.4733	0.6885	0.6786	0.6212	0.2115	0.2650	0.5679	0.6712	0.3243	0.2727			
	SVR	0.3319	0.5633	0.7674	0.6667	0.2759	0.4705	0.6914	0.6804	0.4394	0.0938	0.2805	0.5429	0.6641	0.2297	0.2917			
	Ensemble	0.4618	0.6848	0.7157	0.7143	0.2105	0.5252	0.7267	0.6443	0.5758	0.1915	0.3266	0.5951	0.6425	0.4189	0.3542			
	LSTM	0.3518	0.5748	0.7594	0.7143	0.2105	0.4852	0.6967	0.6708	0.5758	0.1915	0.2866	0.5551	0.6613	0.4189	0.3542			
0.05°	FFNN	0.4233	0.7288	0.7409	0.4921	0.1389	0.6030	0.7951	0.5891	0.8030	0.2090	0.5251	0.7201	0.5395	0.4189	0.1842			
	ANFIS	0.4277	0.7313	0.7381	0.4762	0.1176	0.6113	0.8001	0.5829	0.8030	0.2090	0.5661	0.7258	0.5157	0.6216	0.1481			
	SVR	0.4308	0.7299	0.7361	0.5873	0.1591	0.6217	0.8025	0.5750	0.7273	0.1579	0.5599	0.7316	0.5194	0.5270	0.1875			
	Ensemble	0.5338	0.7537	0.6661	0.6032	0.1364	0.6717	0.8301	0.5357	0.7273	0.1579	0.6536	0.8038	0.4473	0.7568	0.2632			
	LSTM	0.4484	0.7350	0.7180	0.6032	0.1915	0.6256	0.8075	0.5644	0.7576	0.2424	0.5691	0.7563	0.5018	0.6622	0.2344			

As mentioned earlier, 15 sites were employed to further evaluate the proposed downscaling model. Figure 9 displays scatterplots and provides a quantitative comparison of each algorithm, including in-situ SPI data (reference), original PERSIANN-CDR-based SPI, and downscaled SPI. As observed in Fig. 9, the original PERSIANN-CDR's SPI values are spread distant from the bisector line. Furthermore, they tend to exhibit higher differences for extreme values, while other methods performed better in estimating the extremes. Moreover, considering the results of the test step, the ensemble technique led to the most accurate results.

To provide a more detailed comparison, the performance of the ensemble and LSTM methods in generating, for example, SPI-6 compared to the in-situ based values is presented in Table 5.

In Table 5, the presented results indicate that DC scores for 12 stations are higher than 0.5, suggesting that 80 percent of the recorded data could be correctly simulated by the employed methodology. Conversely, DC scores less than 0.5 indicate a lower accuracy, but it does not necessarily imply a failure of the proposed framework in simulating the task. The CC values calculated between in-situ and obtained SPIs, except for Miandoab station, vary from 0.4982 to 0.9613. These outcomes confirm a reliable consistency of SPI indices derived from ground data and the simulated results by the proposed methodology. RMSE scores range from 0.2772 to 0.9513. The FAR and POD scores can examine other aspects of the suggested framework. According to Table 4, the mean POD for LSTM and ensemble methods is 0.77 and 0.81, respectively, showing that about 80% of the drought conditions could be detected by the suggested methodology. Nevertheless, the POD score is low for the Sardasht station (i.e., 0.25). Although the POD score is unacceptable for this site, the suggested framework is still capable of extracting 25% of the drought conditions. Furthermore, the mean FAR is 0.23, indicating that 23% of the extracted drought events are not real.

Comparing the results of Table 5 and Fig. 6, it is clear that in areas with a high correlation between SPI and LST, the accuracy of outputs is higher. However, there is no obvious relation between the high NDVI correlation in Fig. 6 and the results of the models in Table 5. Moreover, considering the detailed outputs presented in Table 5, it is clear that the modeling outcomes for stations 2, 4, 5, 6, 9, 10, and 11 are more accurate than others, while the worst outputs were obtained for stations 13 and 15. Thus it can be concluded that SPI values of stations with a relatively high average precipitation located in the northeast, east, and central regions could be predicted more accurately than in other regions. SPI values of stations with relatively low average precipitation located in the southwest could not be predicted accurately, while the POD value of station 13 is still 60% correct. Taylor diagrams of all SPI indices (SPI-3, SPI-6, and SPI-12) are plotted in

**Fig. 8** Spatial map of SPI-6 in September 2020 from (a) PERSIAN\_CDR data (0.25°), (b) ensemble method (0.05°), (c) LSTM (0.05°); and in June 2021 from (d) PERSIAN\_CDR data (0.25°), (e) ensemble method (0.05°), (f) LSTM (0.05°)

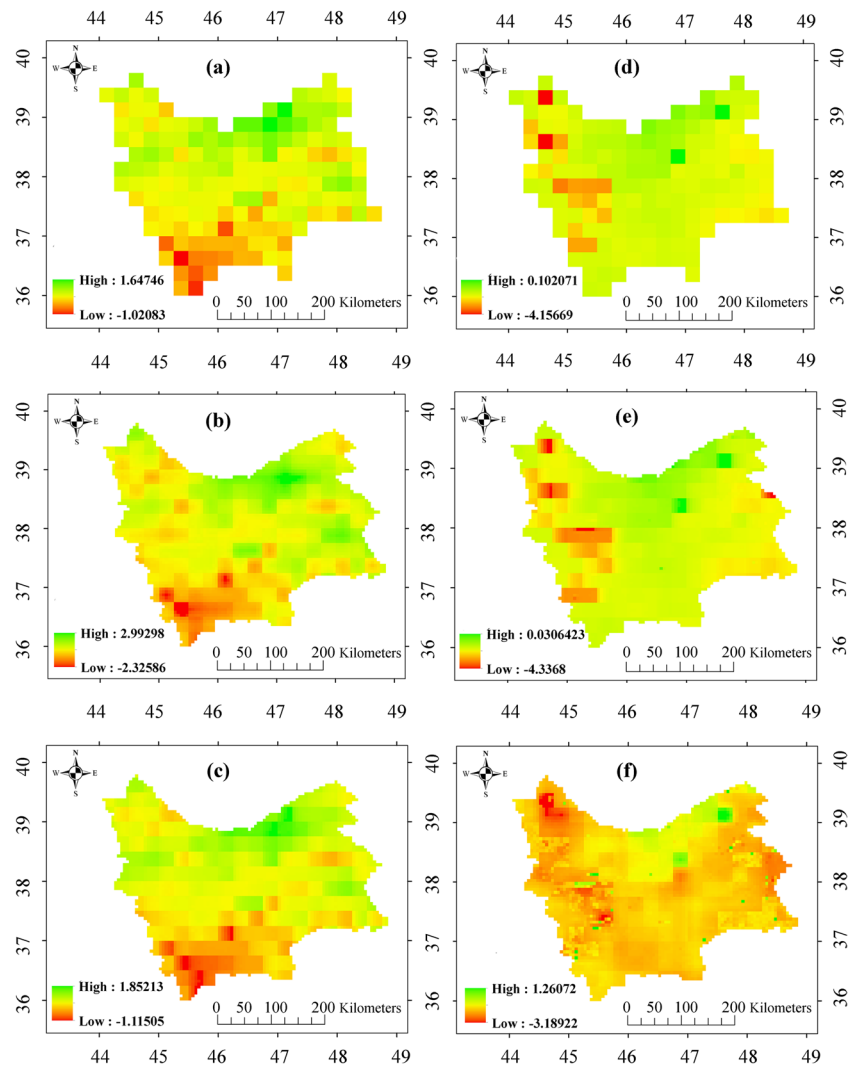
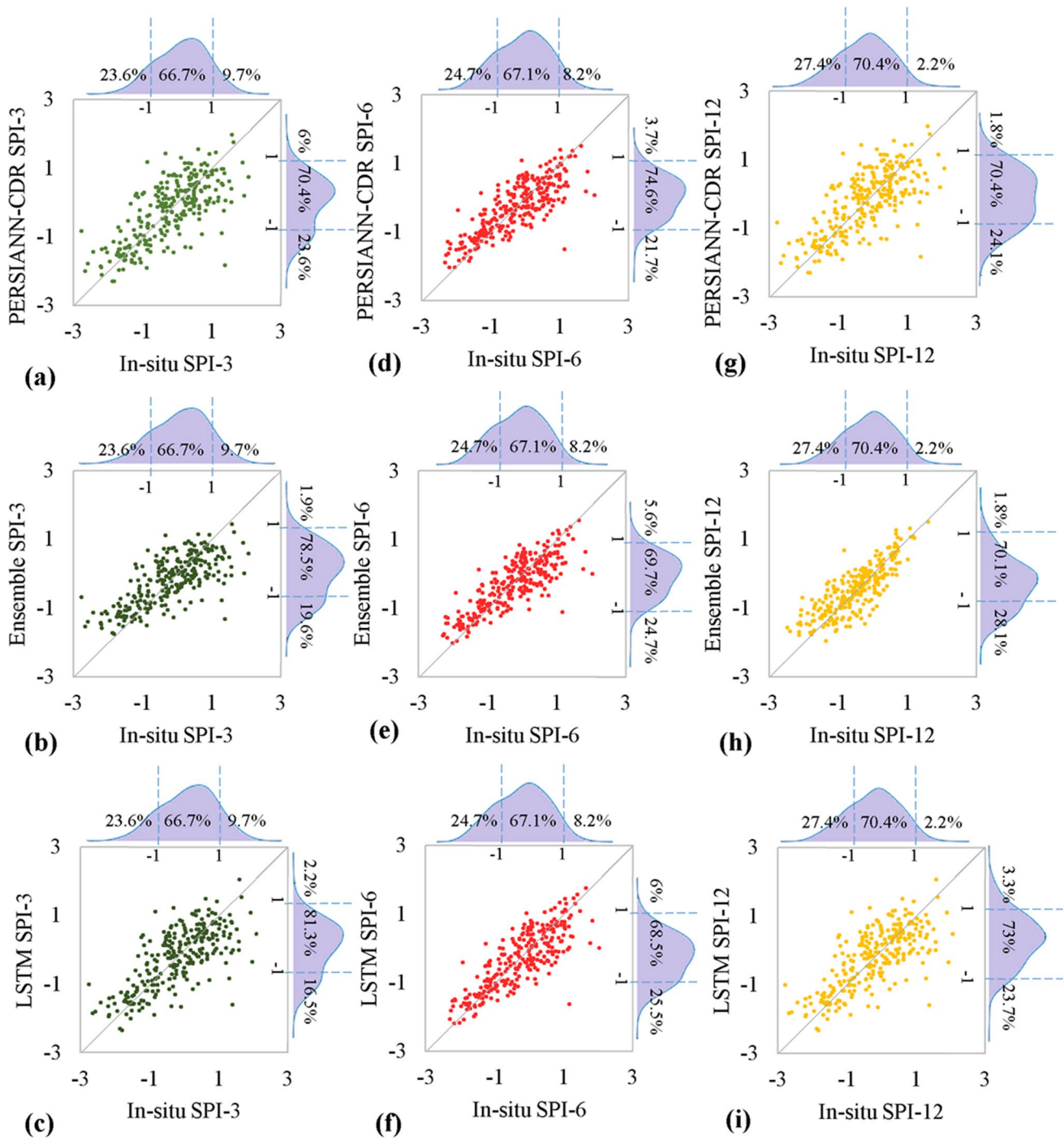


Fig. 10 (Taylor 2001). In a Taylor diagram, smaller distances of the downscaled SPI and reference data (in-situ records) show a better performance of the modeling.

As illustrated in Fig. 10, ensemble predictions are closer to the reference data compared to the other methods. Overall, the results indicate that the ensemble model, with a lower RMSE and higher CC and DC, clearly outperformed the other employed methods. Furthermore, when the available training data is limited, deep learning is not a suitable choice for simulation, and the ensemble strategy can be a suitable alternative.

The proposed approach in the current study for downscaling SPI in the northwest of Iran offers novel contributions and distinctions when compared to existing studies, showcasing advancements in the understanding of AI-driven drought modeling. Similar to Rhee et al. (2010) and Shen et al. (2019), the proposed methodology leverages multiple data sources, incorporating both PERSIAN\_CDR and MODIS-derived variables. However, our distinct

contribution lies in the integration of these diverse datasets to downscale the SPI employing AI models, whereas they focused on monitoring agricultural drought. Aligning with Small et al. (2018), the current study recognizes the significance of temporal scale, specifically highlighting the superior performance of SPI-6 in capturing soil moisture dynamics, emphasizing the understanding of seasonal variations in vegetation response during drought conditions. Furthermore, our study shares similarities with Prodhon et al. (2021) and Shen et al. (2019) in exploring deep learning in a drought context. While others monitor drought, our study uniquely downscale SPI, showcasing the efficiency of ensemble approach for data-limited regions. Acknowledging contextual disparities, our AI-based method overcomes limited data challenges, excelling in improving SPI simulation and detecting 80% of drought conditions. The integration of upscaled SPI as a predictor stands out, distinguishing our work from studies like Mohseni et al. (2021) that lack this element. In contrast to these studies, our ensemble



**Fig. 9** Scatterplots between in-situ and simulated data for SPI-3 obtained by (a) PERSIAN-CDR, (b) Ensemble, (c) LSTM, for SPI-6 obtained by (d) PERSIAN-CDR, (e) Ensemble, (f) LSTM, and for SPI-12 obtained by (g) PERSIAN-CDR, (h) Ensemble, (i) LSTM

post-processing improves accuracy substantially during the test phase. In summary, the current study presents a unique combination of multi-source data integration and upscaled SPI incorporation, contributing to AI-driven drought modeling discourse and highlighting the efficacy of ensemble approaches in regions with data constraints.

### 4 Conclusions

In conclusion, this study aimed to compute high-resolution spatio-temporal SPI indices derived from satellite-based precipitation data, along with thermal and optical parameters, with the ultimate goal of downscaling SPI indices obtained

**Table 5** The performance metrics of ensemble and LSTM methods for downscaled SPI-6 for 15 stations

station No.*	station name	Latitude	Longitude	Mean Precp	Model	DC	CC	RMSE	POD	FAR
1	Maku	39°17'44.24"N	44°30'51.07"E	52.67	Ensemble	0.6331	0.8264	0.5252	0.8	0
					LSTM	0.5938	0.8044	0.5457	0.6	0
2	Germi	39°2'21.69"N	48°4'48.96"E	43.42	Ensemble	0.7551	0.8705	0.4848	1	0
					LSTM	0.7067	0.8036	0.45214	1	0
3	Jolfa	38°55'50.65"N	45°37'43.92"E	46.95	Ensemble	0.5619	0.7587	0.8215	1	0
					LSTM	0.5019	0.7087	0.8759	1	0
4	Kalibar	38°51'53.29"N	47°2'12.9"E	46.39	Ensemble	0.9003	0.9535	0.2829	0.75	0
					LSTM	0.8591	0.9330	0.3241	0.5	0
5	Tabriz	38°5'46.46"N	46°16'25.68"E	44.16	Ensemble	0.7684	0.8882	0.4072	1	0.25
					LSTM	0.7059	0.8237	0.4509	1	0
6	Ardebil	38°15'13.45"N	48°17'59.96"E	47.58	Ensemble	0.7478	0.8678	0.4490	1	0
					LSTM	0.7032	0.8958	0.4065	1	0
7	Sarab	37°56'34.57"N	47°32'18.21"E	46.56	Ensemble	0.6821	0.8353	0.5178	1	0.2
					LSTM	0.6262	0.7957	0.5615	1	0.333
8	Orumiyeh	37°32'59.3"N	45°4'43.06"E	40.94	Ensemble	0.4363	0.8252	0.4264	1	0.667
					LSTM	0.4063	0.7952	0.4338	1	0.667
9	Ajabshir	37°28'43.8"N	45°53'34.54"E	39.69	Ensemble	0.8975	0.9613	0.2772	0.75	0.25
					LSTM	0.8566	0.9141	0.3279	0.75	0.25
10	Khalkhal	37°37'13.95"N	48°31'58.88"E	51.18	Ensemble	0.8714	0.9337	0.3897	0.8	0.333
					LSTM	0.8204	0.8887	0.4605	0.6	0.4
11	Mianeh	37°25'35.16"N	47°43'26.8"E	39.36	Ensemble	0.8569	0.9269	0.3805	0.833	0.167
					LSTM	0.7778	0.8859	0.4741	0.833	0.167
12	Charoymaq	37°7'47.66"N	47°1'28.45"E	36.55	Ensemble	0.6841	0.8661	0.4877	0.833	0.167
					LSTM	0.6356	0.8628	0.5092	0.833	0.167
13	Miandoab	36°55'17.41"N	46°9'45.14"E	35.64	Ensemble	-0.200	0.2644	0.9072	0.6	0.25
					LSTM	-0.320	0.2176	0.9513	0.6	0.4
14	Takab	36°24'18.7"N	47°6'56.19"E	34.15	Ensemble	0.6099	0.8063	0.5277	0.6	0.5
					LSTM	0.5577	0.7800	0.5490	0.6	0.5
15	Sardasht	36°9'30.26"N	45°28'33.56"E	35.64	Ensemble	0.2132	0.5135	0.8124	0.25	0.667
					LSTM	0.1879	0.4982	0.8253	0.25	0.667
Average				42.72	Ensemble	0.6278	0.8065	0.5131	0.8144	0.2301
					LSTM	0.5759	0.7738	0.5432	0.7710	0.2367

\* The location of stations is depicted in Fig. 1

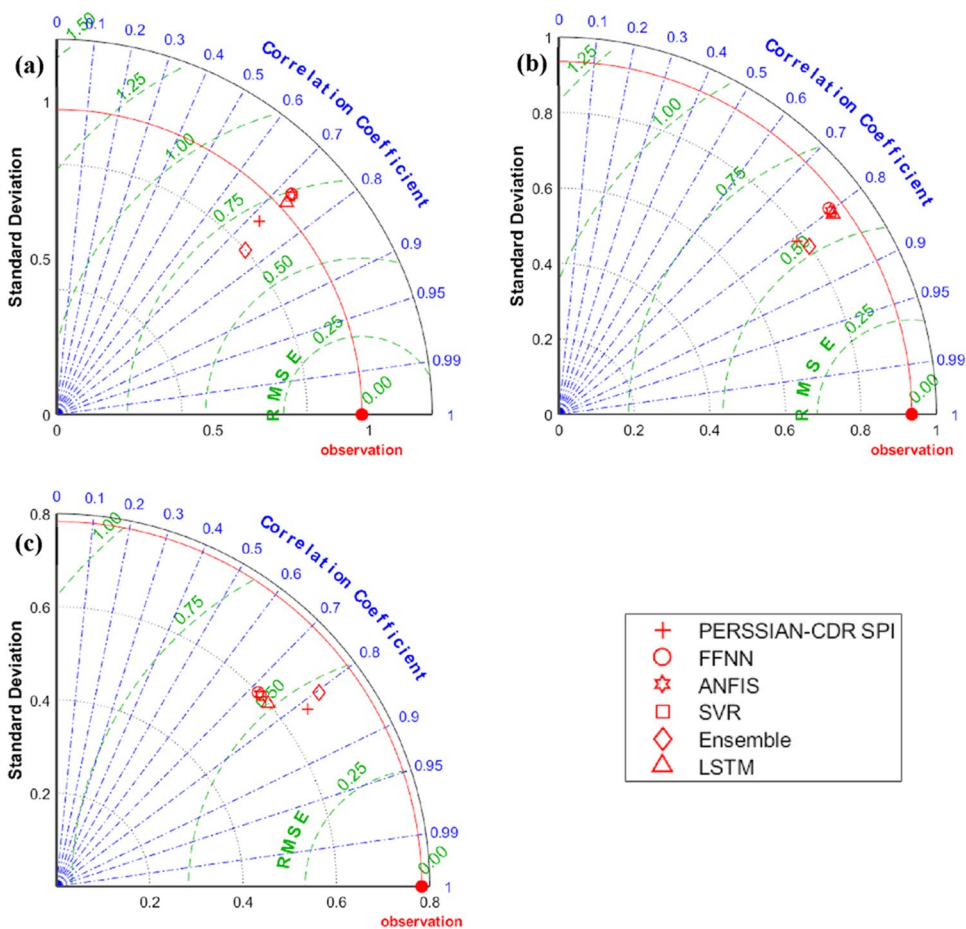
from PERSIANN-CDR products. The proposed framework conducted the downscaling task, considering the correlation between SPI and the vegetation index, along with soil temperature variables derived from MODIS products. The fundamental assumption of the proposed methodology was that drought events exhibit a meaningful relationship with vegetation conditions and soil temperature. On the other hand, the challenges ahead in this research were the inconsistency between the resolutions of different datasets and the application of the trained model on a higher scale to a lower scale. Downscaling procedures involved the utilization of both shallow AI models—specifically, SVR, ANFIS, and FFNN—and the LSTM deep learning method. Following the downscaling process, an ensemble post-processing technique was applied exclusively to the shallow AI models. To

overcome the challenges in the problem, first, the resolution of the high-resolution predictive parameters was reduced according to the existing target resolution at a spatial resolution of  $0.25^\circ$  spanning the years 2000 to 2015; then in this regard, the models were created and trained. In the next step, high-resolution inputs were fed to the trained models to obtain high-resolution output at a spatial resolution of  $0.05^\circ$  for the period from 2016 to 2021 (the goal of the problem at hand).

According to the results obtained, there is no a high similarity between the SPI indices derived from the original PERSIANN-CDR data and the measured ground data. This discrepancy, attributed to spatial resolution differences and errors in satellite measurements, prompted strategic solutions. Through a meticulous examination of the relationship



**Fig. 10** Taylor diagram for (a) SPI-3, (b) SPI-6 and (c) SPI-12 for test phase



between drought-dependent variables and SPI, we successfully downscaled SPI, with modeling SPI-6 demonstrating superior accuracy compared to SPI-3 and SPI-12. This suggests that, over the 6-month period, drought indices are primarily correlated with soil moisture, a key influencing factor on soil temperature and vegetation index. The integration of MODIS LST and NDVI data, coupled with upscaled SPI data, presented an effective downscaling method. It is important to note that precipitation data from different satellites may contain systematic errors stemming from variations in the use of infrared and microwave signals or differences in precipitation modeling, leading to potential inaccuracies. Consequently, precipitation data from alternative satellites may yield better outcomes for specific time periods or regions. Despite potential errors due to precipitation timing around satellite passages, the nonlinear neural ensemble of shallow AI models significantly improved modeling accuracy by up to 20% and 25% in the training and test phases, respectively, achieving the mean DC score of 0.67 in the validation phase. Although the deep learning model, LSTM, showed slightly better performance than shallow AI models during the training phase, its performance was hampered by the limited available data in comparison to

the higher number of parameters involved in deep learning modeling. As a result, the ensemble technique, particularly for unseen data in the test phase, outperformed the deep learning model. Additionally, the proposed methodology successfully detects approximately 80% of drought conditions. A major strength of the proposed framework lies in the pioneering use of upscaled SPI as a predictor for AI models, a novel approach not explored in prior downscaling studies. This consideration, coupled with globally available auxiliary inputs, positions the suggested methodology for broad applicability in various case studies and geographic areas. Additionally, the proposed model holds promise for diverse downscaling purposes, extending beyond satellite products to parameters or global climate model (GCM) data. Despite the reasonable accuracy of SPI indices at the 0.05° scale compared to ground data, it is crucial to acknowledge the inherent uncertainties in the proposed modeling approach, particularly related to the reliance on cloud-free conditions for satellite-based data. While this limitation is not highly significant in the context of the employed monthly modeling scale, it remains an area for consideration in future research. Looking ahead, future studies could benefit from exploring alternative black box models, such as genetic

programming, and validating nonlinear ensemble methods like ANFIS and SVR in place of FFNN. Furthermore, extending the application of the proposed methodology to multi-step ahead modeling could yield valuable insights and enhance its overall applicability.

**Author contribution** All authors have contributed equally to all sections of this manuscript.

**Data availability** The data are available upon reasonable request.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** The authors declare that they have consent to participate.

**Consent for publication** The authors have consented to the publication.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agana NA, Homaifar A (2017) A deep learning based approach for long-term drought prediction. *SoutheastCon 2017*, pp 1–8. <https://doi.org/10.1109/SECON.2017.7925314>
- Anagnostopoulou C (2017) Drought episodes over Greece as simulated by dynamical and statistical downscaling approaches. *Theor Appl Climatol* 129:587–605
- Anagnostopoulou C, Tolika K, Maheras P (2013) Drought index over greece as simulated by a statistical downscaling model. In: Helmis C, Nastos P (eds) *Advances in meteorology, climatology and atmospheric physics*. Springer Atmospheric Sciences, Springer, Berlin, Heidelberg, pp 385–390
- Ashouri H, Hsu KL, Sorooshian S, Braithwaite DK, Knapp KR, Cecil LD, Nelson BR, Prat OP (2015) PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull Am Meteorol Soc* 96:69–83
- Azimi S, Azhdary Moghaddam M (2020) Modeling short term rainfall forecast using neural networks, and Gaussian process classification based on the SPI drought index. *Water Resour Manage* 34:1369–1405
- Baez-Villanueva OM, Zambrano-Bigiarini M, Ribbe L, Nauditt A, Giraldo-Osorio JD, Thinh NX (2018) Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin America. *Atmos Res* 213:34–50
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Crausbay SD, Ramirez AR, Carter SL, Cross MS, Hall KR, Bathke DJ, Betancourt JL, Colt S, Cravens AE, Dalton MS, Dunham JB (2017) Defining ecological drought for the twenty-first century. *Bull Am Meteorol Soc* 98:2543–2550
- Gidey E, Dikinya O, Sebege R, Segosebe E, Zenebe A (2018) Using drought indices to model the statistical relationships between meteorological and agricultural drought in Raya and its environs, Northern Ethiopia. *Earth Syst Environ* 2:265–279
- Goly A, Teegavarapu RS, Mondal A (2014) Development and evaluation of statistical downscaling models for monthly precipitation. *Earth Interact* 18:1–28
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780
- Immerzeel WW, Rutten MM, Droogers P (2009) Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sens Environ* 113:362–370
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685
- Jang JSR, Sun CT, Mizutani E (1997) *Neuro-fuzzy and soft computing—a computational approach to learning and machine intelligence*. Prentice-Hall, New Jersey
- Kazienko P, Lughofer E, Trawiński B (2013) Hybrid and ensemble methods in machine learning. *J Comput Sci* 19:457–461
- Kumar B, Chattopadhyay R, Singh M, Chaudhari N, Kodari K, Barve A (2021) Deep learning–based downscaling of summer monsoon rainfall data over Indian region. *Theor Appl Climatol* 143:1145–1156
- Lee CS, Sohn E, Park JD, Jang JD (2019) Estimation of soil moisture using deep learning based on satellite data: A case study of South Korea. *Gisci Remote Sens* 56:43–67
- Li W, Gao X, Hao Z, Sun R (2022) Using deep learning for precipitation forecasting based on spatio-temporal information: a case study. *Clim Dyn* 58:443–457
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. *Eighth Conf on Applied Climatology*, Anaheim, CA, *Amer Meteor Soc* 17:179–184
- Mohseni F, Mokhtarzade M (2020) A new soil moisture index driven from an adapted long-term temperature-vegetation scatter plot using MODIS data. *J Hydrol* 581:124420
- Mohseni F, Sadr MK, Eslamian S, Areffian A, Khoshfetrat A (2021) Spatial and temporal monitoring of drought conditions using the satellite rainfall estimates and remote sensing optical and thermal measurements. *Adv Space Res* 67:3942–3959
- Neeti N, Murali CA, Chowdary VM, Rao NH, Kesarwani M (2021) Integrated meteorological drought monitoring framework using multi-sensor and multi-temporal earth observation datasets and machine learning algorithms: a case study of central India. *J Hydrol* 601:126638
- Noori R, Karbassi AR, Moghaddamnia A (2011) Assessment of input variables determination on the SVM model performance using PCA, gamma test and forward selection techniques for monthly stream flow prediction. *J Hydrol* 401:177–189
- Nourani V, Behfar N (2021) Multi-station runoff-sediment modeling using seasonal LSTM models. *J Hydrol* 601:126672
- Nourani V, Kheiri A, Behfar N (2022) Multi-station artificial intelligence based ensemble modeling of suspended sediment load. *Water Supply* 22:707–733
- Park S, Im J, Park S, Rhee J (2017) Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. *Agric for Meteorol* 237:257–269
- Prodhan FA, Zhang J, Yao F, Shi L, Pangali Sharma TP, Zhang D, Cao D, Zheng M, Ahmed N, Mohana HP (2021) Deep learning

- for monitoring agricultural drought in South Asia using remote sensing data. *Remote Sens* 13:1715
- Raghavendra NS, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput* 19:372–386
- Rhee J, Im J (2017) Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data. *Agric for Meteorol* 237:105–122
- Rhee J, Im J, Carbone GJ (2010) Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. *Remote Sens Environ* 114:2875–2887
- Sachindra DA, Ahmed K, Rashid MM, Shahid S, Perera BJC (2018) Statistical downscaling of precipitation using machine learning techniques. *Atmos Res* 212:240–258
- Shamseldin AY, O'Connor KM, Liang GC (1997) Methods for combining the outputs of different rainfall-runoff models. *J Hydrol* 197:203–229
- Sharghi E, Nourani V, Behfar N, Tayfur G (2019) Data pre-post processing methods in AI-based modeling of seepage through earthen dams. *Measurement* 147:106820
- Sharghi E, Nourani V, Zhang Y, Ghaneei P (2022) Conjunction of cluster ensemble-model ensemble techniques for spatiotemporal assessment of groundwater depletion in semi-arid plains. *J Hydrol* 610:127984
- Sharifi E, Saghafian B, Steinacker R (2019) Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. *J Geophys Res: Atmos* 124:789–805
- Shen R, Huang A, Li B, Guo J (2019) Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. *Int J Appl Earth Obs Geoinf* 79:48–57
- Small EE, Roesler CJ, Larson KM (2018) Vegetation response to the 2012–2014 California drought from GPS and optical measurements. *Remote Sens* 10:630
- Sun L, Lan Y (2021) Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: localization and comparison with other methods. *Int J Climatol* 41:1128–1147
- Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu KL (2018) A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev Geophys* 56:79–107
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res: Atmos* 106:7183–7192
- Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. *J Hydrol* 330:621–640
- Ullah W, Wang G, Ali G, Tawia Hagan DF, Bhatti AS, Lou D (2019) Comparing multiple precipitation products against in-situ observations over different climate regions of Pakistan. *Remote Sens* 11:628
- Wang F, Tian D, Lowe L, Kalin L, Lehrter J (2021) Deep learning for daily precipitation and temperature downscaling. *Water Resour Res* 57:2020WR029308
- Wang L, Qu JJ (2007) NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophys Res Lett* 34:L20405
- Zhang A, Jia G (2013) Monitoring meteorological drought in semi-arid regions using multi-sensor microwave remote sensing data. *Remote Sens Environ* 134:12–23
- Zhang D, Zhang W, Huang W, Hong Z, Meng L (2017) Upscaling of surface soil moisture using a deep learning model with VIIRS RDR. *ISPRS Int J Geo-Inf* 6:130
- Zhang Y, Li Y, Ji X, Luo X, Li X (2018) Fine-resolution precipitation mapping in a mountainous watershed: geostatistical downscaling of TRMM products based on environmental variables. *Remote Sens* 10:119
- Zhao Z, Di P, Chen SH, Avise J, Kaduwela A, DaMassa J (2020) Assessment of climate change impact over California using dynamical downscaling with a bias correction technique: method validation and analyses of summertime results. *Clim Dyn* 54:3705–3728
- Zhu W, Lv A, Jia S, Sun L (2017) Development and evaluation of the MTVDI for soil moisture monitoring. *J Geophys Res: Atmos* 122:5533–5555

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