ORIGINAL PAPER

Combination of data-driven models and best subset regression for predicting the standardized precipitation index (SPI) at the Upper Godavari Basin in India

Chaitanya B. Pande[1,2] · Romulus Costache[3,4] · Saad Sh. Sammen[5] · Rabeea Noor[6] · Ahmed Elbeltagi[7]

Abstract

Standardized precipitation index prediction and monitoring are essential to mitigating the effect of drought actions on precision farming, environments, climate-smart agriculture, and the water cycle. In this study, four data-driven models, additive regression, random subspace, M5Pruned (M5P), and bagging tree models, were adopted to predict the standardized precipitation index (SPI) at the Upper Godavari Basin for various periods (3 months, 6 months, and 12 months). The data-driven models' input data was pre-processed with machine learning models to increase quality and the model's performance a priori. These four models predicted the SPI-3, SPI-6, and SPI-12 months based on three metrological station data. Based on the statistical performance metrics such as correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), relative absolute error (RAE), and root relative squared error (RRSE), our findings showed that the bagging was the best model for predicting SPI-3 and SPI-6 while the M5P the best for SPI-12 estimation in station 1, while in stations 2 and 3, M5P was superlative in predicting the SPI-3 and SPI-12 months, and the bagging was best in SPI-6. All the best models had acceptable mid-term drought forecasting based on the SPI-3, SPI-6, and SPI-12 months for three stations in the Upper Godavari Basin in India. The machine learning models created in this study produced satisfactory results in short-term and mid-term drought forecasting, and it will be a new strategy for water developers and planners to use for future management and scheduling.

✉ Ahmed Elbeltagi
ahmedelbeltagy81@mans.edu.eg

Chaitanya B. Pande
chaitanay45@gmail.com

Romulus Costache
romuluscostache2000@yahoo.com

1 Indian Institute of Tropical Meteorology, Pune, India

2 Institute of Energy of Infrastructure, Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia

3 Department of Civil Engineering, Transilvania University of Brasov, Brasov, Romania

4 Danube Delta National Institute for Research and Development, 165 Babadag, 820112 Tulcea, Romania

5 Department of Civil Engineering, College of Engineering, University of Diyala, Diyala Governorate, Iraq

6 Department of Agricultural Engineering, Bahuddin Zakariya University, Multan 34200, Pakistan

7 Agricultural Engineering Dept, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

1 Introduction

Drought is one of the climatic hazards that affect large areas of the Earth's surface (Dice and Rodziewicz 2020). It is very clear that when it spreads over a long period of time, this natural phenomenon causes serious economic damage, especially in the field of agriculture (Lopez-Nicolas et al. 2017; Pande et al. 2023b, c, d). Drought poses a serious threat to the economies of many developing countries, especially on the continents of Africa and Asia (Kilimani et al. 2018). Global climate change, which is almost unanimously recognized by the scientific community (Wang et al. 2017), will inevitably lead to an increase in the frequency and duration of the sect in many regions of the globe (Liu et al. 2021). Moreover, the presence of drought can trigger other natural hazards such as forest fires (Aragão et al. 2018). This phenomenon is not only characteristic of arid and hyper-arid areas; drought occurs periodically also throughout other types of ecosystems (Bahrami et al. 2019). Therefore, due to the high importance of the problems that this natural hazard causes around the globe, the study of drought impact on society and, also, of its specific indicators, is a very important

activity in the work of today's researchers (Finn et al. 2018; Tong et al. 2018; Webber et al. 2018).

In the literature, the characterization of drought phenomena is carried out using several indices like the following: Palmer Drought Severity Index (PDSI) (Yang et al. 2020; Yu et al. 2019), Surface Water Supply Index (SWSI) (Duan et al. 2018; Jang et al. 2017), Crop Moisture Index (CMI) (Carrão et al. 2016; Juhasz and Kornfield 1978), Crop Specific Drought Index (CSDI) (Hubbard and Wu 2005; Meyer et al. 1993), Soil Moisture Drought Index (SMDI) (Sohrabi et al. 2015; Xu et al. 2020), Rainfall Anomaly Index (RAI) (Hänsel et al. 2016; Moron 1994), Reclamation Drought Index (RDI) (Weghorst 1996), Effective Precipitation Index (EPI) (Ebrahimpour et al. 2015; Peng-cheng et al. 2016), Bhalme and Mooley Drought Index (BMDI) (Domenikiotis et al. 2004; Ntale and Gan 2003), Effective Drought Index (EDI) (Malik et al. 2021a, b), and Standardized Precipitation Index (SPI) (Wu et al. 2005; Mohamadi et al. 2020).

Of all the previously mentioned indices, one of the most used in the study of climatological drought is SPI. This index is very closely related to soil moisture and also to the groundwater reserve (Spennemann et al. 2015). One of the reasons for the wide use of SPI in various research studies is given by the simplicity of the computation method as well as by the flexibility regarding the various scales of temporal analysis (Choubin et al. 2016). However, the accurate future prediction of SPI values remains a challenge for today's scientific world. In this regard, in order to obtain the highest possible accuracy of the results, so far, the following machine learning models were applied: artificial neural network (Ibrahimi and Baali 2018; Poornima and Pushpalatha 2019; Soh et al. 2018), adaptive neuro-fuzzy inference system (Aghelpour et al. 2020; Ali et al. 2018; Mokhtarzad et al. 2017), support vector regression (Komasi et al. 2018; Roodposhti et al. 2017), support vector machine (Belayneh and Adamowski 2012; Shamshirband et al. 2020), random forest (Yaseen et al. 2021).

In this context, the present research paper aims to (1) enrich the state of knowledge in the field of using machine learning models in SPI prediction, by using the following four algorithms: additive regression model, random subspace, M5P model, and bagging model; (2) select the best developed ML model based on evaluation and validation of the results obtained by a series of statistical indicators such as correlation coefficient, mean absolute error, root mean squared, relative absolute error, and root relative squared error.

2 Methodology

2.1 Study area

The Upper Godavari River basin area is located in the Maharashtra state of India. The river area is covered by 152,199 km^2 supplying approximately 65% of water usage in the state of Selangor. It is located at an elevation of 1067 m, about 80 km from the Arabian Sea. The Timbakeshwar is a source of the Upper Godavari River basin located in the Nashik District of Maharashtra. Ultimately, the river discharges into the Bay of Bengal through a comprehensive tributary network. This river basin is the second biggest river in India. The daily precipitation was observed for each station in the area. This data was collected from the Prediction of Worldwide Energy Resources, NASA. The three weather stations' data were selected and used to combine data-driven models and best subset regression for predicting the standardized precipitation index (SPI). The river basin area is most important for agriculture development, industrial activities, and drinking purposes. This study is based in a river basin of an agricultural area, in which 30 years of daily precipitation time series analysis were carried out to understand climate change by correlating only to the dry spell and wet spell frequencies and some discussions with farmers, without further assessment. Sustainable development to minimize this vastly increasing urban situation is a difficult task to avoid serious implications of environmental deficit in the area, pollution, forest instability, and land-use changes in surface soil cover (Pramudya et al. 2016). The application of a best-suited drought index to the climate prospects noted was considered necessary under those circumstances. For 1989—2019, a sequence of weather data for three stations, including rainfall and temperature, was obtained where only daily data are available from 1989 to 2019. The location map of the Upper Godavari River basin is presented in Fig. 1.

2.2 Standardized precipitation index

The currently created drought index, SPI or SPEI, is reported as an accurate tool for studying and real-time observing the metrological drought situations under heating since numerous scientists and researchers had done SPI drought analysis for the forecasting of future metrological drought events. In this paper, 3, 6, and 12 months of SPI computation were carried out for predicting the standardized precipitation index. SPI was measured based on the daily precipitation data during the years 1989–2019 (30 years of data). The SPI or SPEI calculated value for intensity dryness is such that drought is classified as mild if the SPI or SPEI values vary between 0 and − 1, moderate if from − 1 to − 1.5, severe between − 1.5 and − 2, and extreme when less than − 2. The defined SPI classified is identical to the SPI, because in the calculation, they share a parallel based on the distribution of probabilities (Tan et al. 2015). As per the creation of SPI, three observations situated in various portions of the globe which added various regions such as tropical, monsoon, arid, semi-arid, continental, cold, and oceanic weathers have chosen to create the SPI (Vicente-Serrano et al. 2010). The SPI values were estimated using the SPI package in the R software. Thus, the time scales of SPI implemented in this research are 3, 6, and 12 months. SPI
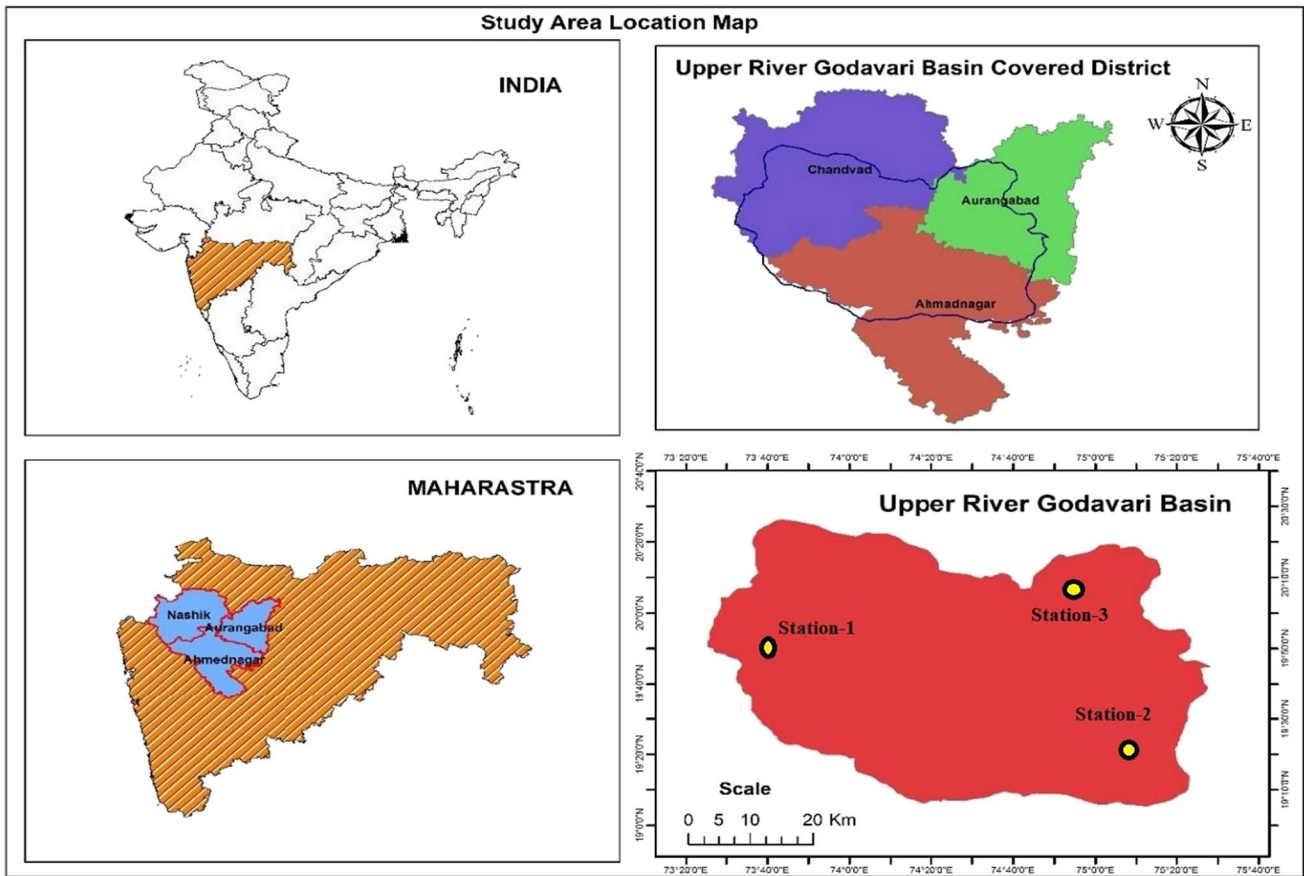
Fig. 1 Location map of the study area

prediction values were computed by using machine learning models. The past SPI from 1989 to 2019 has included additive regression, random subspace, M5P, and bagging models for forecasting the SPI for the test period from 2013 to 2019.

2.3 Machine learning models

2.3.1 Additive regression model

The additive regression model was introduced in the literature by Stone (1985). In this type of model, a dependent variable defined as Y_i ($I = 1, 2,.., n$) represents the sum of many functions that are associated to the following independent variables: $X_{i1}, X_{i2}, \ldots, X_{ip}$. The mathematical relationship on which the additive regression model is based has the following form:

$$Y_i = \sum_{j=1}^{p} f(x_{ij}) + u_i, u_i \sim iid(0, \sigma^2) \quad (1)$$

where $f(x_i)$ represents a nonparametric function, which could be calculated by using a nonparametric regression algorithm. Further, if we consider $E(f) = 0$ ($j = 1,2,\ldots,p$), then the additive regression model equation becomes:

$$E(Y_i | x_{i1}, x_{i2}, \ldots, x_{ip}) = \sum_{j=1}^{p} f(x_{ij}) \quad (2)$$

According to Eq. 2, the additive regression model represents a better version of linear models (Xu and Lin 2015). It should be remarked that the explanatory variables are encoded in a more general form ($f(x_{ij})$) than the initial linear form ($\beta_i x_i$).

2.3.2 Random subspace

Random subspace, which was proposed for the first time by Ho (1998), is an ensemble algorithm that works with a selected subset features of an individual classifier and, finally, using the voting procedure, manages to combine their outputs (Pham et al. 2020). Therefore, the weak individual classifier performance is improved through an ensemble classifier.

Let consider a sample C as a training set of size n, a set $P = (P_1, P_2,\ldots, P_n)$ having as training objects P_i ($i = 1,2,\ldots, n$) which is a q-dimensional vector $Pi = (P_{i1}, P_{i2},\ldots, P_{iq})$ that is characterized by many q features. If we consider $r < q$ features, then r will become a dimensional random subspace

associated with q which is a dimensional feature space. Thus, each object from $Pi = (P_{i1}, P_{i2}, ..., P_{in})$ will be a unit of set sample $P = (P_1, P_2, ..., P_n)$. The random subspace algorithm can be mathematically expressed as follows (Dai et al. 2002):

$$\gamma(s) = argmax \sum_d \delta_{sng}\left(E^d(s)\right), \gamma \tag{3}$$

where $\delta_{i,j}$ represents the Kronecker symbol, while $\gamma = (-1, 1)$ represents a class label associated with the classifier $E^d(s)$ in which $d = 1, 2, ..., D$).

2.3.3 M5P model

The M5P algorithm is a linear tree-based method which is involved in the prediction of continuous variables (Khosravi et al. 2020). Due to the fact that M5P can be characterized by many multivariate linear models, this algorithm has high flexibility (Zhan et al. 2011). The next 3 stages are required to be followed in order to construct the M5P model: (i) tree construction; (ii) tree pruning; (iii) tree smoothing. The growing tree process is intended to maximize the standard deviation reduction (SDR) in order to reach the best performance of the model. The SDR formula can be written as follows (Khosravi et al. 2020):

$$SDR = SD(E) - \sum_i \frac{|E_i|}{|E|} x SD(E_i) \tag{4}$$

where E is a set of cases, E_i represents the ith subset of cases that is obtained following the tree splitting, $SD(E)$ is the standard deviation associated with E, and $SD(E_i)$ represents the standard deviation of E_i.

2.3.4 Bagging model

The bagging model, which was proposed by Breiman (1996), represents an algorithm that consists of a set of basic functions and models which is able to achieve M learners by creating additional data within the training phase (Yariyan et al. 2020). The M training dataset is generated through a random sampling procedure following the substitution of the initial dataset. Within the bagging algorithm, K models are trained with the help of K subsets finally leading to the generation of the final model. The bagging model is a stand-alone one that does not take into account the previous model's precision (Yariyan et al. 2020).

2.3.5 Best combination selection procedure

Feature selection is one of the stages providing a soft computing model to forecast and predict the engineering phenomena when there are many input variables. There are several approaches to specify the best combinations among all possible which are

including best subset regression, mutual information, forward stepwise selection, etc. In the current study, the best subset regression analysis was performed to determine the best input combinations for the SPI model. For this purpose, six statistical criteria, including MSE, determination coefficients (R^2), adjusted (R^2), Mallows' Cp, Akaike's AIC, and Amemiya's PC were computed. The lagged data were prepared as inputs to the models from the 1st (SPI-1) to the 15th (SPI-15). The best subset regression model was applied to select the best input variables in SPI-3-, SPI-6-, and SPI-12-month modeling. It is noteworthy that the total of all datasets were randomly divided into two training and testing subsets. Seventy-five percent of datasets were allocated for training the models and the remaining 25% were considered for validating the developed models.

2.4 Performance metrics for the evaluation of the models

Performance statistics of the correlation coefficient (C.C), mean absolute error (MSE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE) were utilized to measure the applied models of machine learning (Eqs.1 to 5). The following five performance metrics are definite as:

$$C.C = \sqrt{1 - \frac{\sum_{i=1}^{n}(Zi - Z^i)^2}{\sum_{i=1}^{n}(Zi)^2}} \tag{5}$$

where Zi and Z^i is the measures and estimated value; n is the number of value used in the model.

$$MAE = \frac{\sum_{i=1}^{n}\left|(y_i - x_i)\right|}{N} \tag{6}$$

$$RAE = \frac{\sum_{i=1}^{n}|p_i - a_i|}{\sum_{i=1}^{n}|\bar{a} - a_i|} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{N}} \tag{8}$$

$$RRSE = \sqrt{\frac{\sum_{j=1}^{n}(P_{(ij)} - T_j)^2}{\sum_{j=1}^{n}(T_j - \overline{T_{j_j}})^2}} \tag{9}$$

$$d = 1 - \frac{\sum_{j=1}^{N}(P_{ij} - T_j)^2}{\sum_{j=1}^{n}\left(\left|T_j - \overline{P_{ij}}\right| + \left|P_{ij} - \overline{P_{ij}}\right|\right)^2} \tag{10}$$

$$NSE = 1 - \frac{\sum_{j}^{n} (P_{ij} - T_j)^2}{\sum_{j}^{n} (P_{ij} - \overline{P_{ij}})^2}$$ (11)

where $P(ij)$ is the value predicted by the single algorithm i for reported j (out of n data); Tj is the target value for reported j.

3 Results and discussion

In this paper, data of three climate stations are named 1, 2, and 3. The Upper Godavari River basin in India was chosen to develop the SPI index at various scales such as 3, 6, and 12 months. In the study areas, most of the villages have faced problems related to metrological drought conditions and climate parameter changes. Prediction of the SPI drought index data is considered essential for forecasting the metrological drought condition in the study area. Therefore, four machine-learning models including additive regression, random subspace, M5P, and bagging models were adopted for the prediction of the standardized precipitation index for 3 months (SPI-3), 6 months (SPI-6), and 12 months (SPI-12).

3.1 Input selection using the best subset model

For almost every machine-learning model, the input variable selection is essential to obtaining the optimum regression model. There are different techniques and approaches that could be used for input variables selection. One of the commonly used methods is the model-free (filter) based method based on a statistical analysis of model performance. Therefore, several combinations of input variables based on past SPI values (t-1, t-2, t-3,…t-15) were used to predict the SPI. Several statistical indices, which include MSE, R^2, adjusted (R^2), Mallows' (Cp), Akaike's (AIC), Schwarz's (SBC), and Amemiya's (PC), were calculated in order to obtain the best input variables combination.

According to results in Table 1 (A), the best subset regression analysis performance and the finest variables of the SPI-3 model have been observed with seven input variables that include (1st/6th/7th/11th/12th/13th/15th). The best model performance statistics discovered for the SPI-3 model are MSE = 0.540, R^2 = 0.701, adjusted (R^2) = 0.692, Mallows' (Cp) = 2.023, Akaike's (AIC) = − 133.164, Schwarz's (SBC) = − 105.694, and Amemiya's (PC) = 0.318 for station 1 (Table 1). The best subset regression analysis performance and the finest variables of the SPI-6 model have been observed with four input variables include (1st/12th/13th/15th) as shown

the Table 1 (B). The statistics of that model were found as MSE = 0.453, R^2 = 0.802, adjusted (R^2) = 0.799, Mallows' (Cp) = 0.354, Akaike's (AIC) = − 176.295, Schwarz's (SBC) = − 159.127, and Amemiya's (PC) = 0.205 as shown in Table 1 (B). According to the SPI-12 results described in Table 1 (C), the seven variables subset model shows the best results to the finest accuracy of the SPI-12 in station 1. The seven input variables are including (1st/2nd-2/3rd-3/7th/13th/14th/15th) in the SPI-12 model formed an MSE = 0.152, R^2 = 0.944, Adjusted (R^2) = 0.942, Mallows' (Cp) = 2.913, Akaike's (AIC) = − 422.907, Schwarz's (SBC) = − 395.438, and Amemiya's (PC) = 0.059.

Table 2 (A) shows the best subset regression analysis performance and the best input variable combination of the SPI-3 model have been reported in seven variables that include (1st/4th/6th/7th/11th/13th/15th) for station 2. The best model performance results are MSE = 0.528, R^2 = 0.609, adjusted (R^2) = 0.597, Mallows' (Cp) = 4.988, Akaike's (AIC) = − 133.164, Schwarz's (SBC) = − 111.105, and Amemiya's (PC) = 0.415. The best subset regression analysis performance and the finest variables of the SPI-6 model for station 2 have been observed with seven input variables that include (1st/5th/6th/10th/12th/13th/15th). The statistics of that model were found as MSE = 0.478, R^2 = 0.802, Adjusted (R^2) = 0.796, Mallows' (Cp) = 3.313, Akaike's (AIC) = − 161.079, Schwarz's (SBC) = − 133.609, and Amemiya's (PC) = 0.210 as shown in Table 2B. According to the SPI-12 results described in Table 2 (C), the six variables subset model shows the best results to the finest accuracy of the SPI-12 in station 2. These six variables include (1st/7th/12th/13th/14th/15th) in the SPI-12 model formed an MSE = 0.166, R^2 = 0.943, adjusted (R^2) = 0.942, Mallows' (Cp) = 0.200, Akaike's (AIC) = − 404.098, Schwarz's (SBC) = − 380.062, and Amemiya's (PC) = 0.060, as shown in Table 2 (C).

Table 3 (A) shows the best subset regression analysis performance and the best input variables of the SPI-3 model for station 3. The best model has been reported with nine input variables that include (1st/4th/6th/7th/11th/12th/13th/14th/15th). The best model performance results are MSE = 0.525, R^2 = 0.617, adjusted (R^2) = 0.602, Mallows' (Cp) = 5.367, Akaike's (AIC) = − 137.766, Schwarz's (SBC) = − 103.429, and Amemiya's (PC) = 0.414. In Table 3 (B), the best subset regression analysis performance and the finest input variables of the SPI-6 model have been observed with four variables that include (1st/12th/13th/15th). The statistics of that model were found as MSE = 0.455, R^2 = 0.802, adjusted (R^2) = 0.799, Mallows' (Cp) = 0.119, Akaike's (AIC) = − 175.530, Schwarz's (SBC) = − 158.361, and Amemiya's (PC) = 0.205 had the best accuracy in four variables of the SPI-6 model for station 3. According to the SPI-12 results described in Table 3 (C), the fifth

Table 1 The best subset regression analysis for determining the best input combinations to model

(A) SPI-3 for Station 1

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1st-1	0.564	0.680	0.678	5.576	-129.171	-122.304	0.323
2	1st-1 / 11th	0.551	0.688	0.686	1.220	-133.573	-123.272	0.317
3	1st-1 / 13th / 15th	0.548	0.691	0.687	1.158	-133.689	-119.954	0.317
4	1st-1 / 11th / 13th / 15th	0.545	0.694	0.689	0.868	-134.061	-116.892	0.316
5	1st-1 / 11th / 12th / 13th / 15th	0.543	0.697	0.690	0.896	-134.125	-113.522	0.316
6	1st-1 / 6th / 11th / 12th / 13th / 15th	0.542	0.699	0.691	1.547	-133.546	-109.510	0.317
7	**1st-1 / 6th / 7th / 11th / 12th / 13th / 15th**	**0.540**	**0.701**	**0.692**	**2.023**	**-133.164**	**-105.694**	**0.318**
8	1st-1 / 4th-4 / 6th / 7th / 11th / 12th / 13th / 15th	0.541	0.702	0.691	3.479	-131.743	-100.840	0.320
9	1st-1 / 4th-4 / 6th / 7th / 11th / 12th / 13th / 14th / 15th	0.543	0.703	0.690	5.035	-130.218	-95.881	0.322
10	1st-1 / 4th-4 / 5th-5 / 6th / 7th / 11th / 12th / 13th / 14th / 15th	0.545	0.703	0.689	6.825	-128.444	-90.673	0.324
11	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 6th / 7th / 11th / 12th / 13th / 14th / 15th	0.546	0.703	0.688	8.482	-126.811	-85.607	0.326
12	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 7th / 11th / 12th / 13th / 14th / 15th	0.548	0.704	0.687	10.308	-124.998	-80.359	0.329
13	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 7th / 8th / 11th / 12th / 13th / 14th / 15th	0.551	0.704	0.686	12.221	-123.091	-75.018	0.332
14	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 7th / 8th / 9th / 11th / 12th / 13th / 14th / 15th	0.553	0.704	0.684	14.119	-121.200	-69.695	0.335
15	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 7th / 8th / 9th / 10 th / 11th / 12th / 13th / 14th / 15th	0.555	0.704	0.683	16.000	-119.328	-64.389	0.337

(B) SPI-6 for Station 1

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1st-1	0.462	0.796	0.795	1.674	-174.778	-167.911	0.206
2	1st-1 / 15th	0.462	0.797	0.795	2.757	-173.706	-163.405	0.207
3	1st-1 / 13th / 15th	0.455	0.801	0.798	0.185	-176.392	-162.657	0.204
4	**1st-1 / 12th / 13th / 15th**	**0.453**	**0.802**	**0.799**	**0.354**	**-176.295**	**-159.127**	**0.205**
5	1st-1 / 11th / 12th / 13th / 15th	0.454	0.803	0.799	1.585	-175.099	-154.497	0.206
6	1st-1 / 5th-5 / 6th / 12th / 13th / 15th	0.454	0.804	0.798	2.914	-173.804	-149.768	0.207
7	1st-1 / 5th-5 / 6th / 7th / 12th / 13th / 15th	0.454	0.805	0.799	3.624	-173.163	-145.693	0.207
8	1st-1 / 5th-5 / 6th / 7th / 11th / 12th / 13th / 15th	0.454	0.806	0.799	4.652	-172.193	-141.289	0.208
9	1st-1 / 2nd-2 / 5th-5 / 6th / 7th / 11th / 12th / 13th / 15th	0.454	0.806	0.798	6.032	-170.853	-136.515	0.209
10	1st-1 / 2nd-2 / 3rd-3 / 5th-5 / 6th / 7th / 11th / 12th / 13th / 15th	0.455	0.807	0.798	7.350	-169.580	-131.809	0.211
11	1st-1 / 2nd-2 / 3rd-3 / 5th-5 / 6th / 7th / 11th / 12th / 13th / 14th / 15th	0.455	0.808	0.798	8.526	-168.461	-127.257	0.212
12	1st-1 / 2nd-2 / 3rd-3 / 5th-5 / 6th / 7th / 8th / 11th / 12th / 13th / 14th / 15th	0.457	0.808	0.797	10.175	-166.838	-122.200	0.213
13	1st-1 / 2nd-2 / 3rd-3 / 5th-5 / 6th / 7th / 8th / 10th / 11th / 12th / 13th / 14th / 15th	0.459	0.808	0.797	12.081	-164.940	-116.867	0.215
14	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 7th / 8th / 10th / 11th / 12th / 13th / 14th / 15th	0.461	0.808	0.796	14.009	-163.017	-111.511	0.217
15	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 7th / 8th / 9th / 10th / 11th / 12th / 13th / 14th / 15th	0.463	0.808	0.795	16.000	-161.026	-106.087	0.219

(C) SPI-12 for Station 1

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1st-1	0.183	0.931	0.931	40.597	-387.480	-380.612	0.069
2	1st-1 / 7th	0.164	0.939	0.938	14.132	-411.440	-401.138	0.063
3	1st-1 / 13th / 14th	0.155	0.942	0.941	2.999	-422.488	-408.753	0.060
4	1st-1 / 7th / 13th / 14th	0.153	0.943	0.942	0.008	-425.648	-408.479	0.059
5	1st-1 / 7th / 13th / 14th / 15th	0.152	0.944	0.942	0.894	-424.815	-404.213	0.059
6	1st-1 / 6th / 13th / 14th / 15th	0.152	0.944	0.942	0.894	-424.815	-404.213	0.059
7	**1st-1 / 2nd-2 / 3rd-3 / 7th / 13th / 14th / 15th**	**0.152**	**0.944**	**0.942**	**2.913**	**-422.907**	**-395.438**	**0.059**
8	1st-1 / 2nd-2 / 3rd-3 / 6th / 13th / 14th / 15th	0.152	0.944	0.942	2.913	-422.907	-395.438	0.059
9	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 6th / 13th / 14th / 15th	0.153	0.944	0.942	4.137	-421.732	-390.828	0.060
10	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 6th / 11th / 13th / 14th / 15th	0.153	0.944	0.942	5.361	-420.559	-386.222	0.060
11	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 11th / 13th / 14th / 15th	0.153	0.944	0.942	7.216	-418.714	-380.943	0.061
12	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 9th / 11th / 13th / 14th / 15th	0.154	0.944	0.942	9.182	-416.750	-375.546	0.061
13	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 9th / 10th / 11th / 13th / 14th / 15th	0.155	0.944	0.941	11.068	-414.872	-370.234	0.062
14	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 8th / 9th / 10th / 11th / 13th / 14th / 15th	0.155	0.945	0.941	13.003	-412.941	-364.869	0.062
15	1st-1 / 2nd-2 / 3rd-3 / 4th-4 / 5th-5 / 6th / 8th / 9th / 10th / 11th / 12th / 13th / 14th / 15th	0.156	0.945	0.941	15.000	-410.945	-359.439	0.063

The best model for the selected selection criterion is displayed in blue

input variables model shows the best results to the finest accuracy of the SPI-12 model in station 3. The five variables such as (1st/6th/12th/13th/15th) model formed an MSE = 0.150, R^2 = 0.944, adjusted (R^2) = 0.943, Mallows' (Cp) = − 1.078, Akaike's (AIC) = − 428.007, Schwarz's (SBC) = − 407.404, and Amemiya's (PC) = 0.058; all these performance metrics indicated the five variables best input combination for SPI-12 model for station 3 (Table 3 (C)).

3.2 Evaluation of machine learning models

In this paper, the SPI drought index for different time scales including 3,6, and 12 months for the Upper Godawari River basin in India was predicted through four machine learning models, namely additive regression, random subspace, M5P, and the bagging. Different input combinations from three meteorological stations were

Table 2 The best subset regression analysis for determining the best input combinations to model

(A) SPI-3 for station 2

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1TH	0.565	0.570	0.569	14.722	-128.811	-121.943	0.433
2	1TH / 11th	0.535	0.595	0.591	3.022	-140.287	-129.986	0.412
3	1TH / 6TH / 11th	0.535	0.597	0.591	4.091	-139.232	-125.497	0.414
4	1TH / 11th / 14th / 15th	0.532	0.601	0.593	3.881	-139.492	-122.324	0.414
5	1TH / 10th / 11th / 14th / 15th	0.532	0.603	0.594	4.710	-138.698	-118.096	0.415
6	1TH / 4TH / 6TH / 11th / 14th / 15th	0.529	0.607	0.596	4.506	-138.986	-114.950	0.414
7	1TH / 4TH / 6TH / 7TH / 11th / 13th / 15th	0.528	0.609	**0.597**	4.988	-138.575	-111.105	0.415
8	1TH / 4TH / 6TH / 7TH / 11th / 12th / 13th / 15th	0.528	0.611	0.597	6.114	-137.495	-106.592	0.417
9	1TH / 4TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th	0.528	0.613	0.597	7.075	-136.593	-102.256	0.419
10	1TH / 3TH / 4TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th	0.528	0.614	0.596	8.347	-135.366	-97.595	0.421
11	1TH / 3TH / 4TH / 6TH / 7TH / 9TH / 11th / 12th / 13th / 14th / 15th	0.529	0.616	0.596	9.433	-134.340	-93.136	0.423
12	1TH / 2TH / 3TH / 4TH / 6TH / 7TH / 9TH / 11th / 12th / 13th / 14th / 15th	0.529	0.617	0.596	10.626	-133.204	-88.566	0.425
13	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 9TH / 11th / 12th / 13th / 14th / 15th	0.530	0.618	0.595	12.206	-131.655	-83.583	0.428
14	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.533	0.618	0.593	14.065	-129.806	-78.300	0.431
15	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.535	0.618	0.591	16.000	-127.876	-72.936	0.435

(B) SPI-6 for station 2

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1TH	0.485	0.794	0.793	0.494	-163.560	-156.692	0.208
2	1TH / 15th	0.486	0.795	0.793	1.620	-162.449	-152.148	0.209
3	1TH / 13th / 15th	0.479	0.799	0.796	-0.495	-164.683	-150.948	0.207
4	1TH / 12th / 13th / 15th	0.479	0.799	0.796	0.548	-163.679	-146.511	0.208
5	1TH / 10th / 12th / 13th / 15th	0.478	0.801	0.796	1.347	-162.936	-142.334	0.208
6	1TH / 6TH / 10th / 12th / 13th / 15th	0.478	0.801	0.796	2.368	-161.965	-137.928	0.209
7	**1TH / 5TH / 6TH / 10th / 12th / 13th / 15th**	**0.478**	**0.802**	**0.796**	**3.313**	**-161.079**	**-133.609**	**0.210**
8	1TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 15th	0.478	0.803	0.796	4.343	-160.107	-129.204	0.211
9	1TH / 2TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 15th	0.479	0.804	0.796	5.883	-158.597	-124.260	0.212
10	1TH / 2TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th	0.480	0.804	0.795	7.372	-157.143	-119.372	0.214
11	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th	0.481	0.805	0.795	8.859	-155.692	-114.487	0.215
12	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 8TH / 11th / 12th / 13th / 14th / 15th	0.483	0.805	0.794	10.543	-154.030	-109.391	0.217
13	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 8TH / 10th / 11th / 12th / 13th / 14th / 15th	0.484	0.805	0.793	12.139	-152.464	-104.392	0.218
14	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 10th / 11th / 12th / 13th / 14th / 15th	0.486	0.805	0.793	14.026	-150.585	-99.079	0.220
15	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.489	0.805	0.792	16.000	-148.613	-93.674	0.222

The best model for the selected selection criterion is displayed in blue

(C)SPI-12 for station 2

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1TH	0.199	0.930	0.930	39.044	-367.257	-360.389	0.070
2	1TH / 7TH	0.180	0.937	0.937	14.376	-389.638	-379.337	0.064
3	1TH / 12th / 13th	0.171	0.941	0.940	2.990	-400.931	-387.196	0.061
4	1TH / 6TH / 12th / 13th	0.168	0.942	0.941	0.913	-403.137	-385.968	0.060
5	1TH / 7TH / 12th / 13th / 15th	0.166	0.943	0.942	-0.803	-405.039	-384.437	0.060
6	**1TH / 7TH / 12th / 13th / 14th / 15th**	**0.166**	**0.943**	**0.942**	**0.200**	**-404.098**	**-380.062**	**0.060**
7	1TH / 6TH / 7TH / 12th / 13th / 14th / 15th	0.167	0.943	0.941	1.734	-402.594	-375.125	0.060
8	1TH / 3TH / 6TH / 7TH / 12th / 13th / 14th / 15th	0.167	0.943	0.941	3.235	-401.127	-370.224	0.061
9	1TH / 3TH / 6TH / 7TH / 10th / 12th / 13th / 14th / 15th	0.167	0.943	0.941	4.950	-399.432	-365.095	0.061
10	1TH / 3TH / 4TH / 5TH / 6TH / 7TH / 12th / 13th / 14th / 15th	0.168	0.944	0.941	6.644	-397.759	-359.988	0.062
11	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 12th / 13th / 14th / 15th	0.169	0.944	0.941	8.313	-396.114	-354.910	0.062
12	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 10th / 12th / 13th / 14th / 15th	0.169	0.944	0.941	10.077	-394.368	-349.730	0.062
13	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 10th / 12th / 13th / 14th / 15th	0.170	0.944	0.940	12.037	-392.411	-344.339	0.063
14	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 10th / 11th / 12th / 13th / 14th / 15th	0.171	0.944	0.940	14.010	-390.440	-338.935	0.064
15	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.171	0.944	0.940	16.000	-388.451	-333.512	0.064

The best model for the selected selection criterion is displayed in blue

used and the best model (with the best input combination) was adopted according to the statistical index analysis. The value of past SPI (t-1, t-2, t-3,…t-15) was used as input variables in order to predicate the value of future SPI. Meteorological data of 20 years from 2000 to 2019 have been collected and used to build the predictive models. The performance of the machine learning models was evaluated by calculating the arithmetical indices, viz. C.C, NSE, IW, MAE, RMSE, RAE (%), and RRSE (%). The predictive models given were repeatedly performed in order to maintain steady and reliable results.

Table 3 The best subset regression analysis for determining the best input combinations to model

(A) SPI-3 for station 3

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1TH	0.552	0.583	0.581	8.802	-133.893	-127.025	0.421
2	1TH / 11th	0.529	0.602	0.599	0.014	-142.711	-132.410	0.405
3	1TH / 6TH / 11th	0.529	0.604	0.599	0.918	-141.839	-128.104	0.406
4	1TH / 6TH / 7TH / 11th	0.527	0.607	0.600	1.136	-141.685	-124.516	0.407
5	1TH / 4TH / 6TH / 7TH / 11th	0.528	0.609	0.600	2.313	-140.542	-119.940	0.409
6	1TH / 6TH / 7TH / 11th / 14th / 15th	0.526	0.612	0.601	2.536	-140.405	-116.369	0.409
7	1TH / 4TH / 6TH / 7TH / 11th / 14th / 15th	0.525	0.614	0.602	3.344	-139.664	-112.194	0.410
8	1TH / 4TH / 6TH / 7TH / 11th / 12th / 13th / 15th	0.525	0.615	0.601	4.498	-138.561	-107.658	0.412
9	**1TH / 4TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th**	**0.525**	**0.617**	**0.602**	**5.367**	**-137.766**	**-103.429**	**0.414**
10	1TH / 4TH / 6TH / 7TH / 10th / 11th / 12th / 13th / 14th / 15th	0.527	0.618	0.601	6.992	-136.168	-98.397	0.417
11	1TH / 2TH / 3TH / 4TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th	0.528	0.619	0.599	8.693	-134.487	-93.283	0.420
12	1TH / 2TH / 3TH / 4TH / 6TH / 7TH / 10th / 11th / 12th / 13th / 14th / 15th	0.530	0.619	0.598	10.290	-132.919	-88.281	0.423
13	1TH / 2TH / 3TH / 4TH / 6TH / 7TH / 8TH / 9TH / 11th / 12th / 13th / 14th / 15th	0.532	0.620	0.597	12.206	-131.010	-82.938	0.426
14	1TH / 2TH / 3TH / 4TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.534	0.620	0.595	14.035	-129.194	-77.688	0.429
15	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.536	0.620	0.593	16.000	-127.231	-72.292	0.433

(B) SPI-6 for station 3

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1TH	0.463	0.796	0.795	1.240	-174.205	-167.338	0.206
2	1TH / 15th	0.463	0.797	0.795	2.313	-173.146	-162.845	0.207
3	1TH / 13th / 15th	0.456	0.801	0.798	-0.140	-175.717	-161.982	0.205
4	**1TH / 12th / 13th / 15th**	**0.455**	**0.802**	**0.799**	**0.119**	**-175.530**	**-158.361**	**0.205**
5	1TH / 11th / 12th / 13th / 15th	0.455	0.803	0.798	1.490	-174.188	-153.586	0.206
6	1TH / 5TH / 6TH / 12th / 13th / 15th	0.456	0.804	0.798	2.648	-173.071	-149.035	0.207
7	1TH / 5TH / 6TH / 7TH / 12th / 13th / 15th	0.455	0.805	0.798	3.397	-172.392	-144.922	0.208
8	1TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 15th	0.455	0.805	0.798	4.544	-171.296	-140.393	0.209
9	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 12th / 13th / 15th	0.456	0.806	0.798	6.008	-169.867	-135.530	0.210
10	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 15th	0.457	0.807	0.798	7.261	-168.664	-130.893	0.211
11	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 11th / 12th / 13th / 14th / 15th	0.458	0.807	0.797	8.620	-167.349	-126.145	0.212
12	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 8TH / 11th / 12th / 13th / 14th / 15th	0.459	0.807	0.797	10.324	-165.667	-121.029	0.214
13	1TH / 2TH / 3TH / 5TH / 6TH / 7TH / 8TH / 10th / 11th / 12th / 13th / 14th / 15th	0.461	0.808	0.796	12.135	-163.870	-115.798	0.215
14	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 10th / 11th / 12th / 13th / 14th / 15th	0.463	0.808	0.795	14.025	-161.988	-110.483	0.217
15	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.465	0.808	0.794	16.000	-160.015	-105.076	0.219

The best model for the selected selection criterion is displayed in blue

(C) SPI-12 for station 3

No. of variables	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
1	1TH	0.183	0.931	0.931	41.894	-387.474	-380.607	0.069
2	1TH / 6TH	0.164	0.939	0.938	15.290	-411.434	-401.133	0.063
3	1TH / 12th / 13th	0.155	0.942	0.941	4.092	-422.482	-408.747	0.060
4	1TH / 6TH / 12th / 13th	0.153	0.943	0.942	1.077	-425.642	-408.473	0.059
5	**1TH / 6TH / 12th / 13th / 15th**	**0.150**	**0.944**	**0.943**	**-1.078**	**-428.007**	**-407.404**	**0.058**
6	1TH / 6TH / 10th / 12th / 13th / 15th	0.151	0.944	0.943	0.455	-426.502	-402.466	0.059
7	1TH / 2TH / 3TH / 6TH / 12th / 13th / 15th	0.151	0.945	0.943	1.806	-425.193	-397.724	0.059
8	1TH / 2TH / 3TH / 4TH / 6TH / 12th / 13th / 15th	0.151	0.945	0.943	3.035	-424.016	-393.113	0.059
9	1TH / 2TH / 3TH / 4TH / 6TH / 12th / 13th / 15th	0.151	0.945	0.943	4.500	-422.590	-388.253	0.060
10	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 10th / 12th / 13th / 15th	0.152	0.945	0.942	6.317	-420.786	-383.015	0.060
11	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 10th / 12th / 13th / 14th / 15th	0.153	0.945	0.942	8.200	-418.912	-377.707	0.061
12	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 8TH / 10th / 12th / 13th / 14th / 15th	0.153	0.945	0.942	10.121	-416.996	-372.358	0.061
13	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 8TH / 9TH / 10th / 12th / 13th / 14th / 15th	0.154	0.945	0.942	12.073	-415.048	-366.976	0.062
14	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 12th / 13th / 14th / 15th	0.155	0.945	0.941	14.025	-413.100	-361.595	0.062
15	1TH / 2TH / 3TH / 4TH / 5TH / 6TH / 7TH / 8TH / 9TH / 10th / 11th / 12th / 13th / 14th / 15th	0.155	0.945	0.941	16.000	-411.127	-356.187	0.063

The best model for the selected selection criterion is displayed in blue

The best model for the selected selection criterion is displayed in blue

3.2.1 Evaluation of SPI for station 11

The statistical analysis of the performance of predictive models for the testing datasets for station 11 is given in Table 4. The results showed that M5P (NSE = 0.64–0.95 and RMSE = 0.36–0.97) and Bagging model (NSE = 0.66–0.93 and RMSE = 0.4–0.95) were found as the best models for SPI prediction. The lowest performance models for drought index

(SPI) prediction for all time scales and for different stations were found when the additive regression and random subsurface are adopted. Recently, the suitability and ability of the M5T technique are improved by several other studies such as Sattari and Sureh (2019).

In addition, the scatter plots between the observed and predicted (SPIs, SPI6 and SPI12) for the testing models of station 1 are shown Figs. 2, 3, and 4. The results indicate that the performance of M5T and bagging predictive models have a high correlation with the observation while the additive regression and random subsurface predictive models have the lowest correlation with observed SPI especially in SPI-3 and SPI-6. The best values of the correlation coefficient of the predictive models for SPI-12 are found ($R^2 = 0.95$) when the M5T predictive model was used, while the lowest values of the correlation coefficient are found ($R^2 = 0.52$) when the additive regression predictive model was used. However, there was no significant difference between the results obtained with M5T and Bagging predictive models (Table 4).

In order to evaluate the uncertainty in SPI prediction for the Upper Godawari River basin, a box plot was used as shown in Fig. 5. The box plot includes the first-quarter, second-quarter and third-quarter values of all the predictive models and observed SPI. It is clear from Fig. 5 how the M5T represents the best predictive model for SPI prediction followed by the bagging predictive model. In addition, it is found that the fluctuations of the additive regression and random subsurface predictive models were far from the range of the observed SPI. Hence, it could be concluded that the M5T is more suitable for the prediction of SPI with different time scales in station 1.

3.2.2 Evaluation of SPI for station 2

The statistical analysis of the performance of predictive models for the testing datasets for station 1 is given in Table 5. As in the results of station 1, the M5P model (NSE = 0.69–0.97 and RMSE = 0.27–0.72) and the bagging model (NSE = 0.72–0.94 and RMSE = 0.4–0.6) were found as the best models for SPI prediction for station no. 2. In addition, the lowest performance models for all time scales have been found when the additive regression and random subsurface are adopted. Recently, the suitability and ability of the M5T technique are improved by several other studies such as (Sattari and Sureh 2019).

The graphical evaluation using scatter plots between the observed and predicted (SPIs, SP-I6, and SPI-12) for the testing models of station 2 are shown in Figs. 6, 7, and 8). The results indicate that the performance of M5T and bagging predictive models have a high correlation with the observation while the additive regression and random subsurface predictive models have the lowest correlation with observed SPI especially in SPI-3 and SPI-6. The best values of the correlation coefficient of the predictive models for SPI-12 are found ($R^2 = 0.98$) when the M5T predictive model was used for SPI-12 estimation, while the lowest values of the correlation coefficient are found ($R^2 = 0.47$) when the additive regression predictive model was used for SPI-6 estimation. However, the performance of additive regression and random subsurface predictive models for the estimation of SPI-6 was better than their performance for SPI-3 estimation.

Figure 9 presents the box plot for the predicted and observed values of SPI for station 2. It is clear from Fig. 9 the M5T represents the best predictive model for SPI prediction compared with the other models followed by the bagging predictive model. In addition, it is found that the fluctuations of the additive regression and random subsurface predictive models were far from the range of the observed SPI. Hence, it could be concluded that the additive regression and random subsurface predictive models

Models	C.C	MAE	RMSE	RAE (%)	RRSE (%)	NSE	d
SPI-3							
Additive regression	0.72	0.95	1.29	77.07	77.88	0.42	0.67
Random subspace	0.73	0.91	1.24	74.18	74.5	0.64	0.87
M5P	0.81	0.75	0.97	60.86	58.56	0.64	0.87
Bagging	0.81	0.76	0.95	61.69	57.11	0.66	0.88
SPI-6							
Additive regression	0.83	0.72	1.04	75.94	63.79	0.75	0.92
Random subspace	0.92	0.71	1.04	75.94	63.79	0.59	0.8
M5P	0.92	0.45	0.67	47.57	41.04	0.83	0.95
Bagging	0.93	0.42	0.65	44.72	39.63	0.84	0.95
SPI-12							
Additive regression	0.93	0.36	0.68	41.05	43.95	0.81	0.96
Random subspace	**0.95**	**0.27**	**0.51**	**30.10**	**32.15**	**0.9**	**0.97**
M5P	**0.98**	**0.21**	**0.36**	**23.86**	**23.34**	**0.95**	**0.99**
Bagging	**0.97**	**0.24**	**0.4**	**27.34**	**25.66**	**0.93**	**0.98**

Table 4 Statistical analysis of model performance for predicting SPI-3, SPI-6, and SPI-12 in station 11
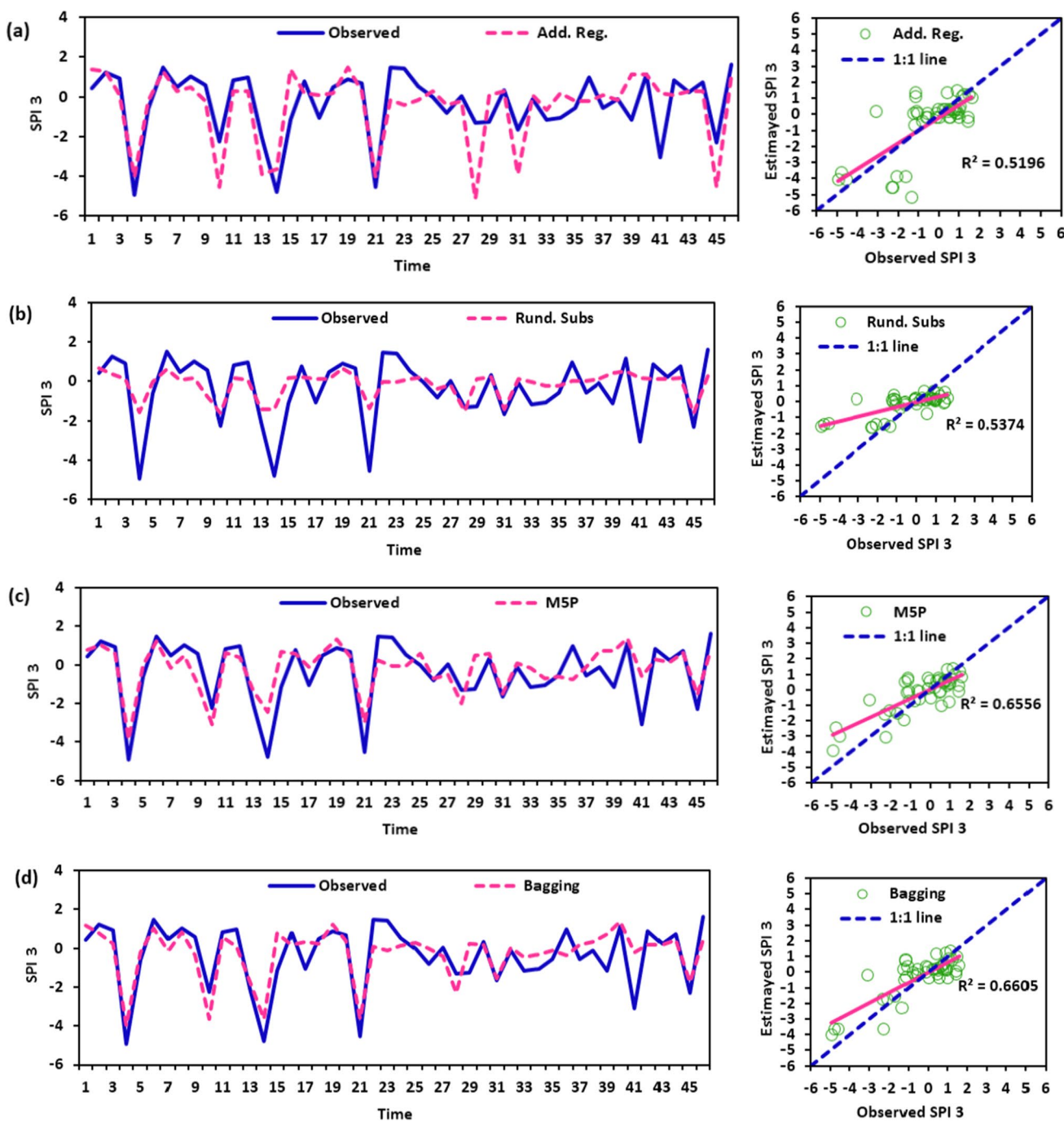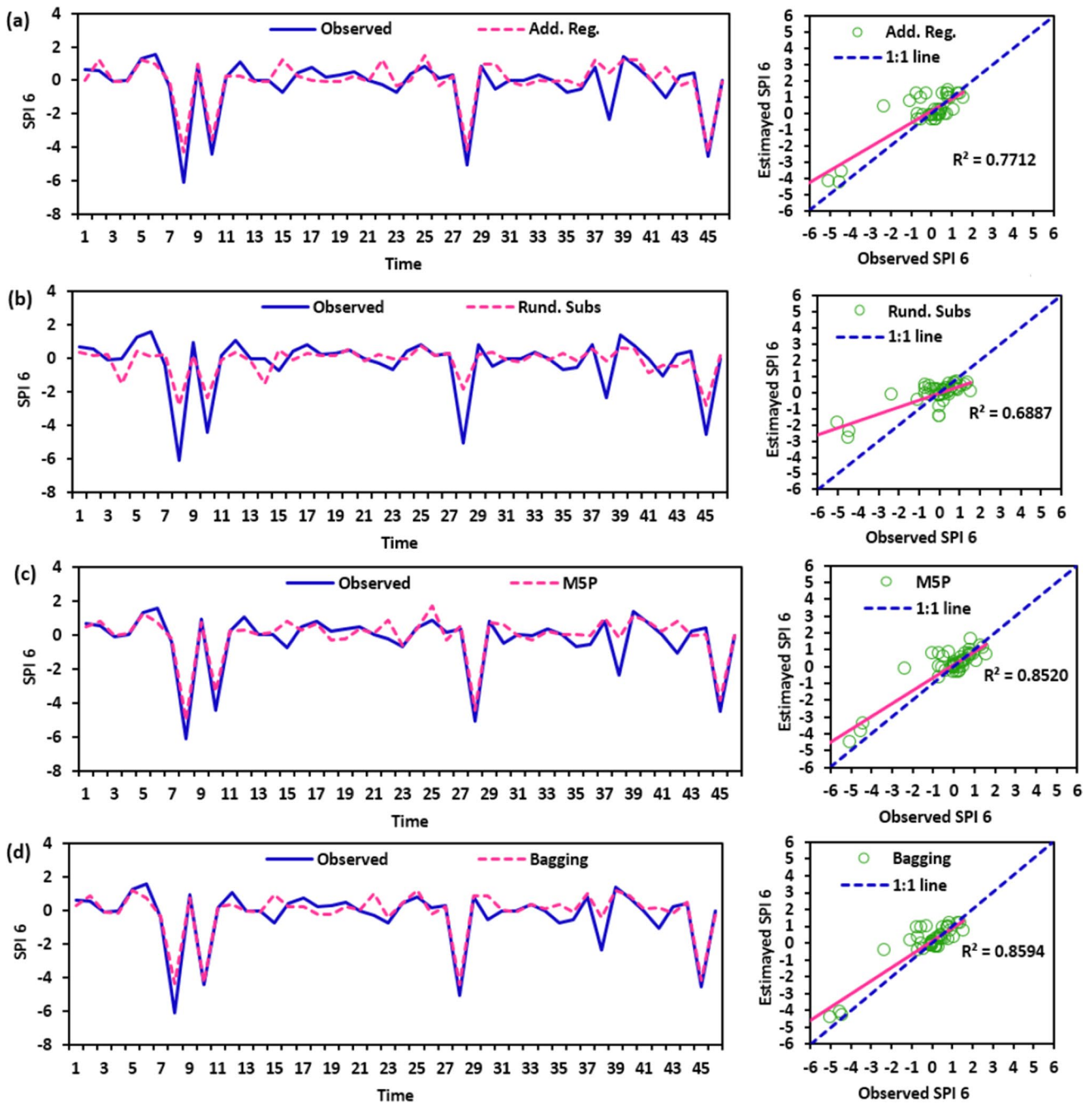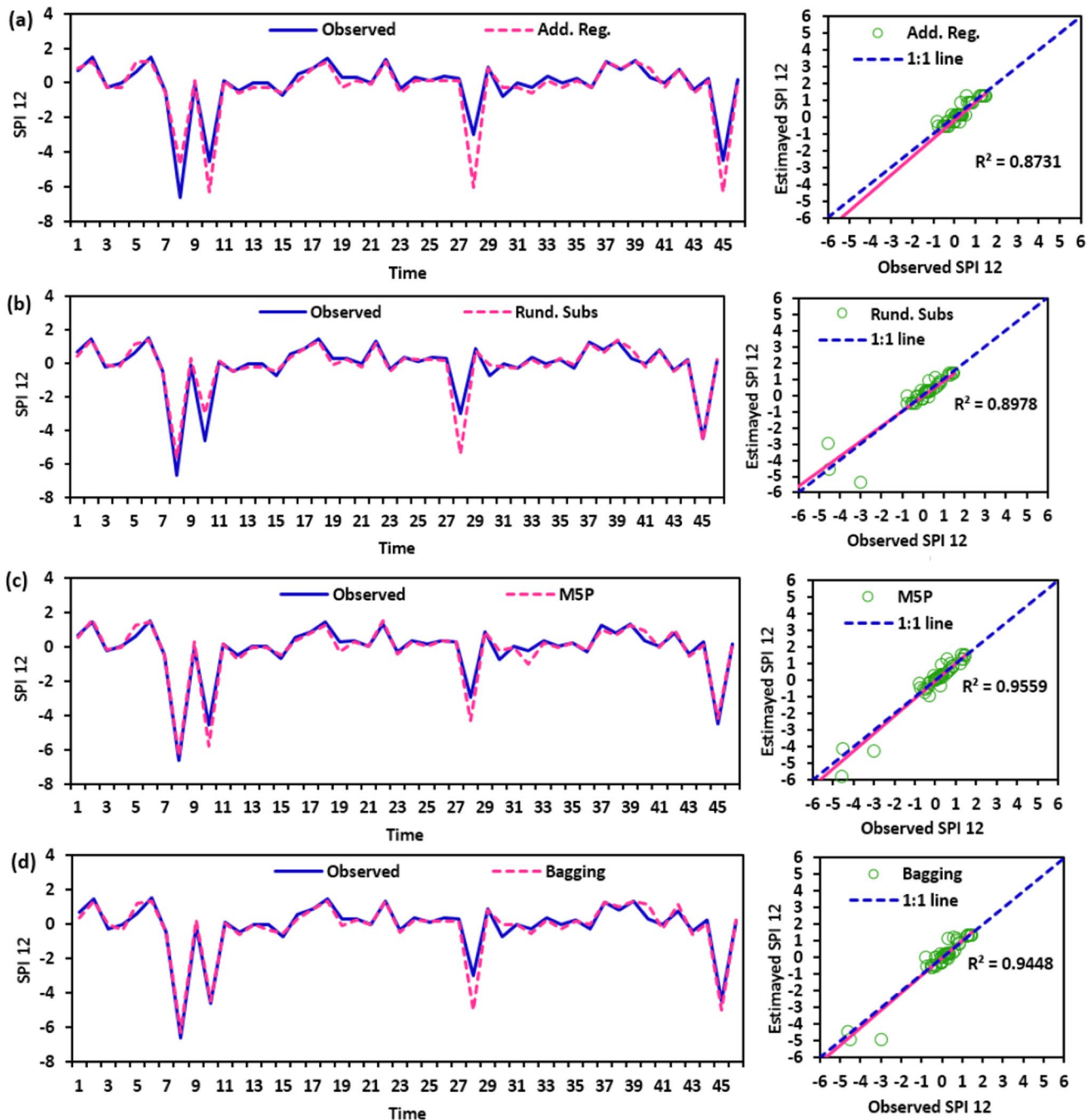
Fig. 2 Observed SPI-3 versus estimated SPI-3. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 1

are not suitable for the prediction of SPI with different time scales in station 2.

3.2.3 Evaluation of SPI for station 3

The statistical analysis of the performance of predictive models for the testing datasets for station 3 is given in Table 6.

Similar to the results of station 1 and station 2, the results showed that the M5P predictive model (NSE = 0.69–0.97 and RMSE = 0.28–0.66) and Bagging model (NSE = 0.60–0.92 and RMSE = 0.45–0.7) were found as the best models for SPI prediction for station 3. The lowest performance models for drought index (SPI) prediction for all time scales and for different stations was found when the additive regression and

Fig. 3 Observed SPI-6 versus Estimated SPI-6 **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 1

random subsurface are adopted. Recently, the suitability and ability of the M5T technique are improved by several other studies such as (Sattari and Sureh 2019).

As for station 1 and station 2, the scatter plots between the observed and predicted (SPIs, SPI6, and SPI12) for the testing models of station 3 are shown in Figs. 10, 11, and 12). The results indicate that the performance of M5T and bagging predictive models have a high correlation with the observation while the additive regression

and random subsurface predictive models have the lowest correlation with observed SPI especially in SPI-3 and SPI-6. The best values of the correlation coefficient are found ($R^2 = 0.98$) when the M5T predictive model was used for SPI-12 estimation, while the lowest values of the correlation coefficient are found ($R^2 = 0.45$) when the additive regression predictive model was used. However, there was no significant difference between the results obtained with M5T and bagging predictive models for
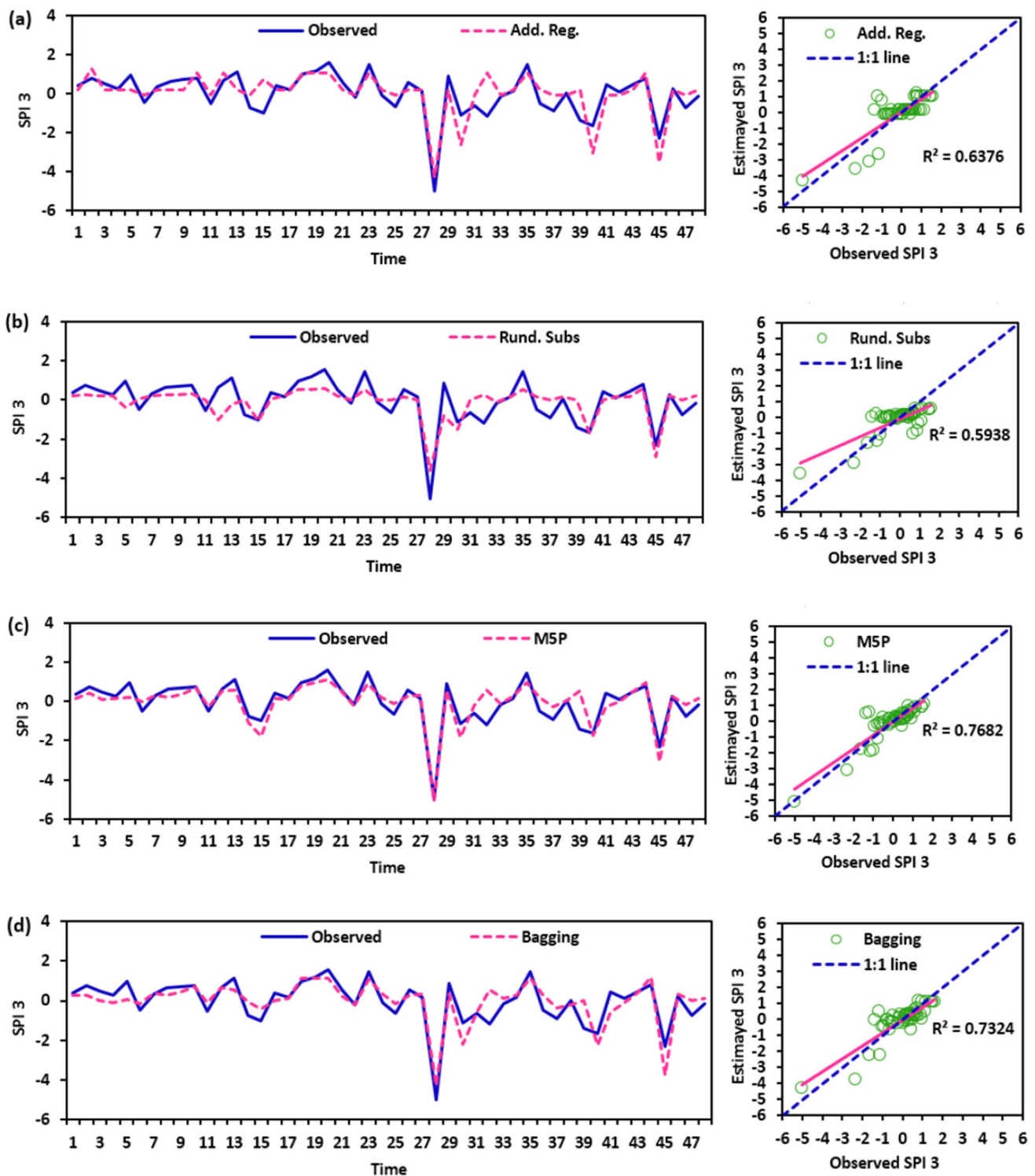
Fig. 4 Observed SPI-12 versus Estimated SPI-12. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 1

SPI-6 estimation. Figure 13 presents the box plot for the predicted and observed values of SPI for station 3. It is clear from Fig. 13 the M5T represents the best predictive model for SPI prediction compared with the other models followed by the Bagging predictive model. In addition, it is found that the fluctuations of the additive regression and random subsurface predictive models were far from the range of the observed SPI. Hence, it could be concluded that the additive regression and random subsurface predictive models are not suitable for the prediction of SPI with different time scales in station 3 and M5T is considered suitable for SPI prediction. Overall, the results revealed that all of the machine learning techniques used in this study could predicate the SPI with a high time scale (SPI-12) with acceptable accuracy and this conclusion is agree with that one improved (Yaseen et al. 2021).
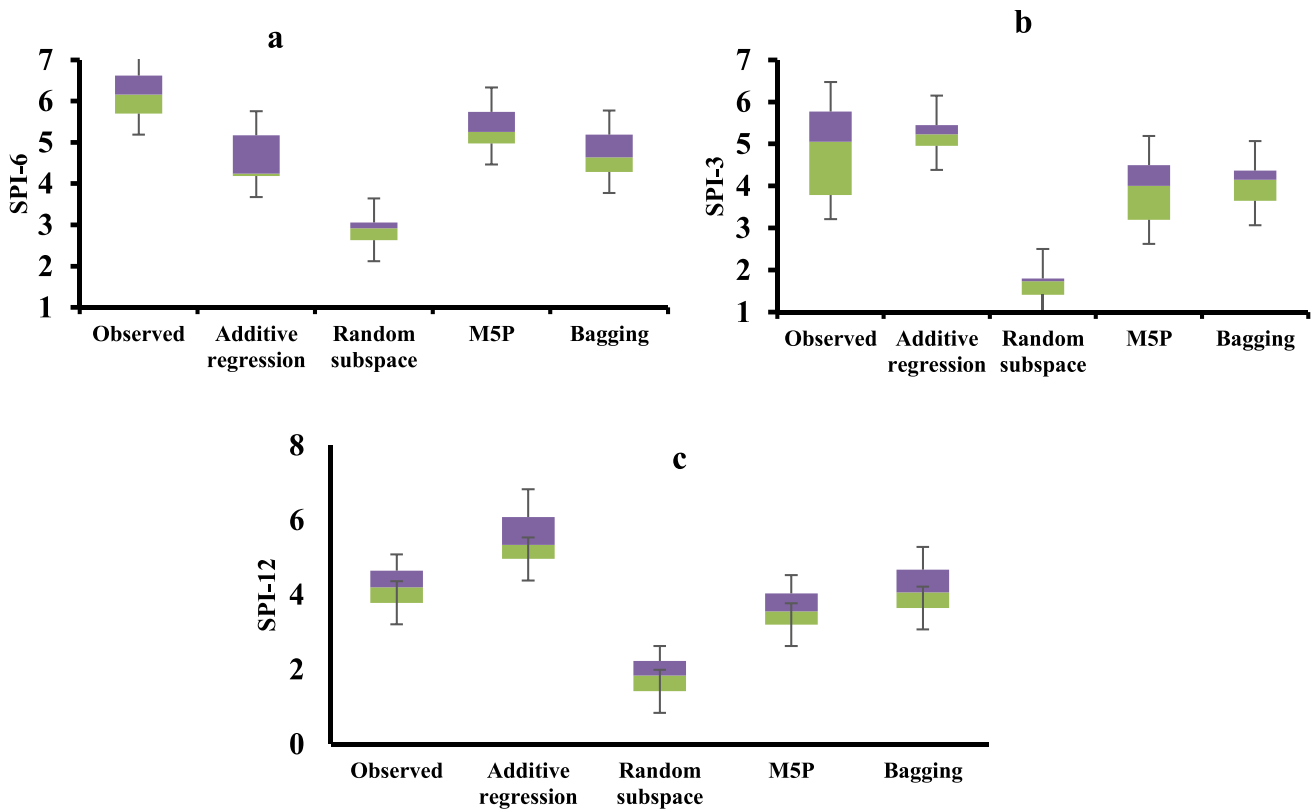
Fig. 5 Box plot presentation of the performance of the predictive models. **a** SPI-3. **b** SPI-6. **c** SPI-12 for station 1

Table 5 Statistical analysis of model performance for predicting SPI-3, SPI-6, and SPI-12 in station 2

Models	C.C	MAE	RMSE	RAE (%)	RRSE (%)	NSE	d
SPI-3							
Additive regression	0.8	0.52	0.71	64.71	63.9	0.59	0.89
Random subspace	0.77	0.55	0.7191	68.13	64.5	0.58	0.85
M5P	0.88	0.38	0.55	47.5	48.94	0.76	0.93
Bagging	0.86	0.45	0.59	55.34	52.63	0.72	0.92
SPI-6							
Additive regression	0.67	0.62	1.1	76.85	84.62	0.29	0.58
Random subspace	0.82	0.6	0.83	72.25	63.77	0.59	0.81
M5P	0.84	0.48	0.72	59.98	55.73	0.69	0.91
Bagging	0.89	0.41	0.6	50.86	45.8	0.79	0.94
SPI-12							
Additive regression	0.97	0.38	0.56	37.11	34.45	0.88	0.97
Random subspace	0.95	0.39	0.53	42.57	32.17	0.9	0.97
M5P	0.99	0.17	0.27	18.79	16.17	0.97	0.99
Bagging	0.98	0.26	0.4	28.64	24.64	0.94	0.98

4 Discussion

In India, droughts regularly have an impact on farming and farmers' lives. To lessen the effects of drought in the area (Orimoloye et al. 2022), reliable drought prediction is crucial. The majority of meteorological stations in India lack the dependable rainfall and temperature data required for drought research and prediction over longer time periods (Shelar et al. 2022; Kumar Gautam et al. 2022). In order to get beyond the restrictions of the climatic data, ML techniques were utilized in this work (Elbeltagi et al. 2023a). The current study showed that ML models can
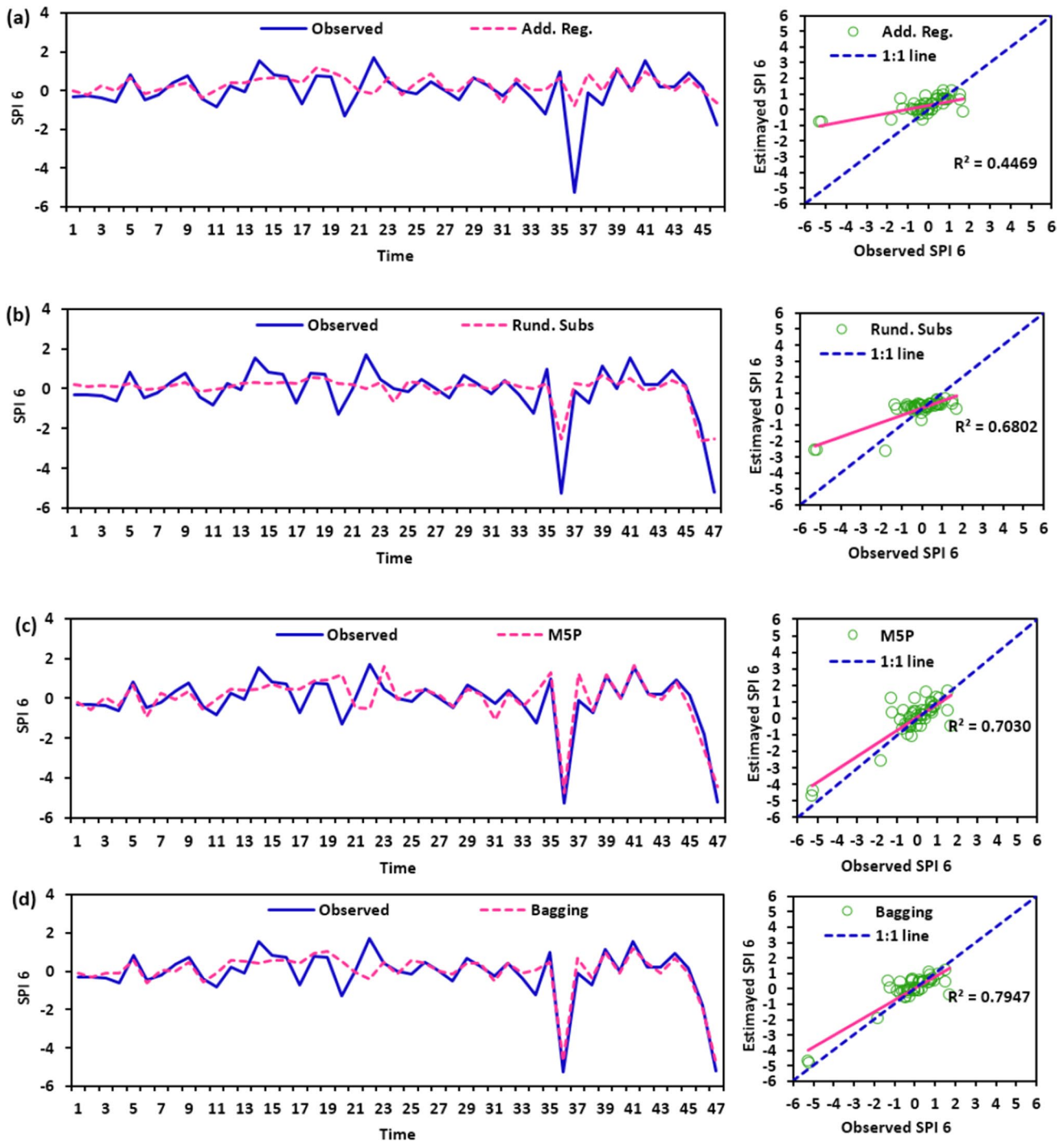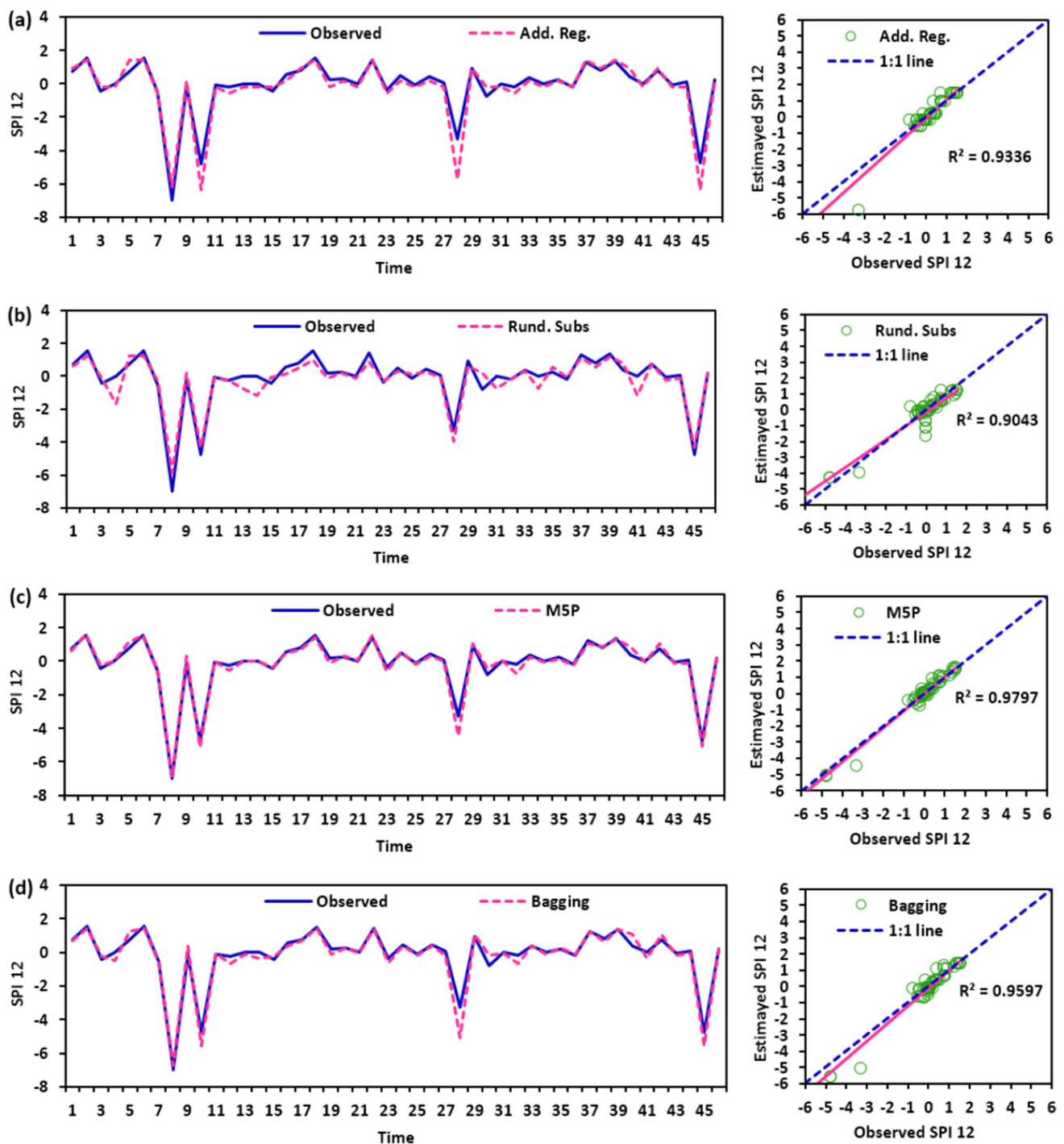
Fig. 6 Observed SPI-3 versus Estimated SPI-3. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 2

anticipate the SPI, the most popular DI, accurately over a multi-month horizon (i.e., 3, 6, and 12). The best subset regression analysis was used to optimize the SPI-3, SPI-6, and SPI-12-month models. Based on the statistical performance metrics, research results showed that the whole best models i.e., (Bagging and M5P) had acceptable forecasting of the mid-term drought forecasting based on the SPI-3, SPI-6, and SPI-12 months for three stations

Fig. 7 Observed SPI-6 versus Estimated SPI-6. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 2

in the Upper Godavari Basin in India. Different regions of India might duplicate the temporal variability of SPIs. This model could help decision-makers and experts in the water sector make wise choices (Pande et al. 2022, 2023a).

For the cases of mid-term dryness, the examined models more accurately predicted SPIs. Our results were contrasted with those of recent research carried out in other places, including Bangladesh, Ethiopia, India, and Iran. When training and testing durations were taken into account, the investigated models more accurately predicted SPIs for mid-term drought circumstances. These results support the research by (Malik et al. 2021a; Yaseen et al. 2021). Additionally,
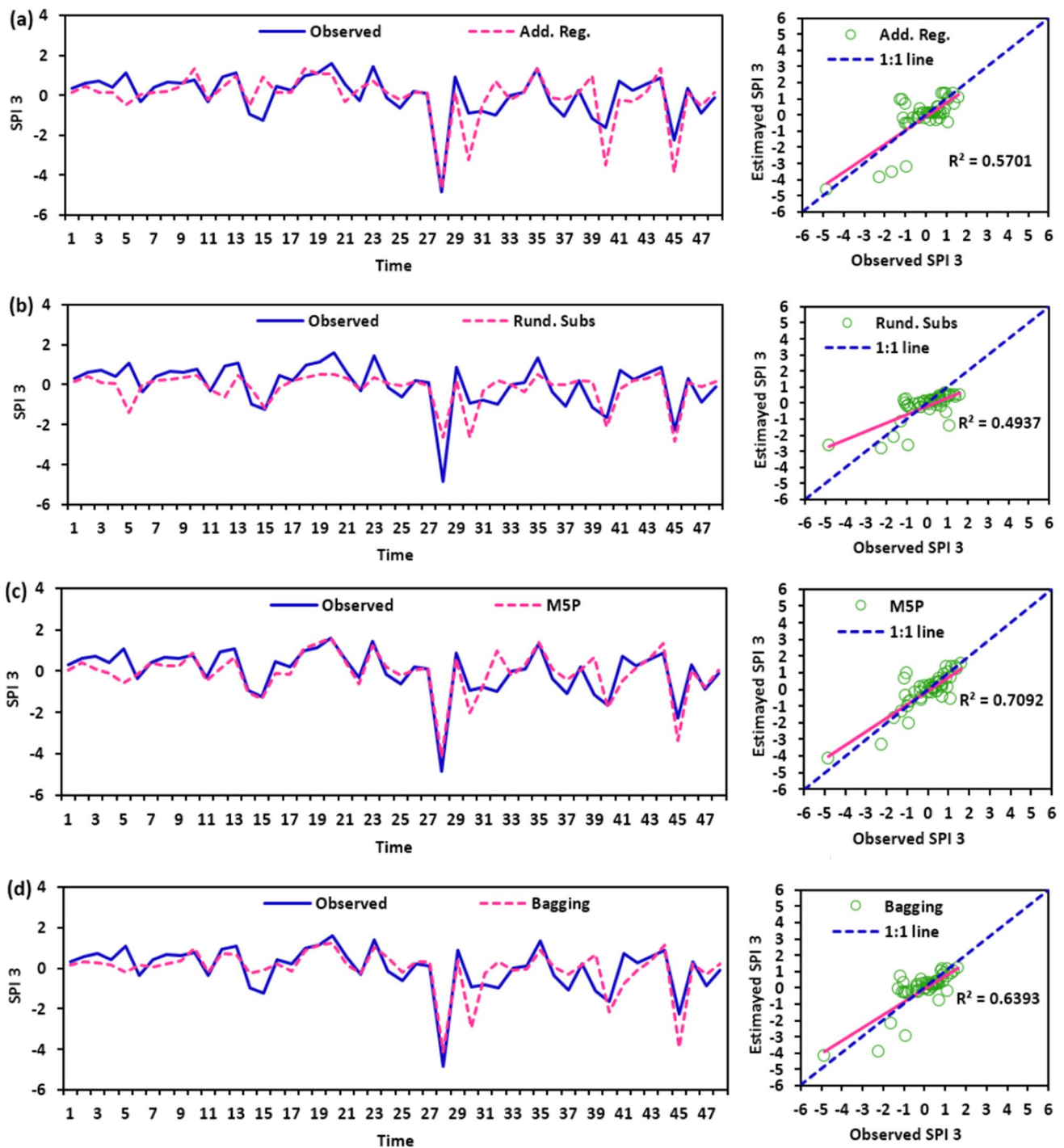
Fig. 8 Observed SPI-12 versus Estimated SPI-12. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 2

monsoon months are more prone to severe drought than other times of the year, with June exhibiting the greatest vulnerability. Serious droughts are more likely to occur in September. The bagging model proved to be superior among the chosen models during the training and testing phases for each timeline of SPI (i.e., SPI-3, SPI-6, and SPI-12). It

agrees with the findings of Ditthakit et al. (2021). In Bangladesh, Yaseen et al. (2021) looked at the effectiveness of machine learning (ML) techniques such random forest (RF), bagging, M5P Tree, extreme learning machine (ELM), and online sequential-ELM (OSELM) in predicting (SPI) at 4-month horizons (i.e., 1, 3, and 12). According to the study,
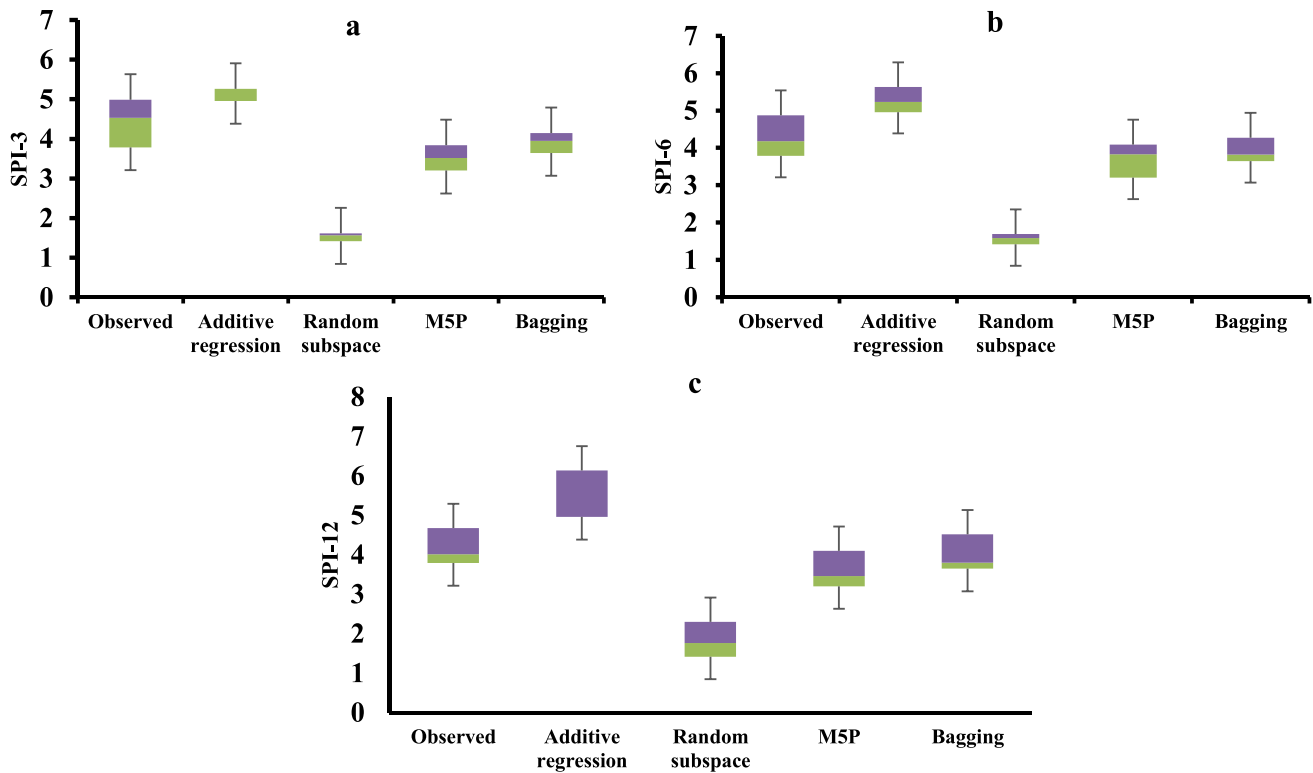
Fig. 9 Box plot presentation of the performance of the predictive models. **a** SPI-3. **b** SPI-6. **c** SPI-12 for station 2

Table 6 Statistical analysis of model performance for predicting SPI-3, SPI-6, and SPI-12 in station 3

Models	C.C	MAE	RMSE	RAE (%)	RRSE (%)	NSE	d
SPI-3							
Additive regression	0.76	0.58	0.84	70.26	75.7	0.42	0.86
Random subspace	0.70	0.59	0.8	71.23	72.8	0.47	0.77
M5P	0.84	0.42	0.62	50.44	55.82	0.69	0.91
Bagging	0.8	0.51	0.7	62.04	63.08	0.6	0.89
SPI-6							
Additive regression	0.82	0.51	0.74	66.58	59.2	0.65	0.9
Random subspace	0.67	0.63	0.93	82.14	74.2	0.45	0.75
M5P	0.86	0.46	0.66	60.1	52.63	0.72	0.92
Bagging	0.88	0.43	0.6	55.9	48.02	0.77	0.93
SPI-12							
Additive regression	0.96	0.35	0.37	39.54	36.88	0.86	0.97
Random subspace	0.96	0.34	0.45	38.47	29.2	0.91	0.98
M5P	0.99	0.18	0.28	19.68	17.91	0.97	0.99
Bagging	0.97	0.27	0.45	29.83	0.29	0.92	0.98

bagging and M5P provided the most accurate predictions for the 3-, 6-, and 12-month SPI. Three machine learning techniques-artificial neural networks (ANNs), support vector regression (SVR), and M5P-were used by Belayneh et al. (2016). They came to the conclusion that M5P provided the superior model performance for SPI-3 (3-month SPI) and SPI-6 (6-month SPI) forecasting multi-scale drought index (Pande et al. 2022, 2023a).

This is anticipated since the smoothness and unpredictability of the SPI time series get worse as the time scale gets longer. As was discovered in the current work, more linear data enhances the performance of machine learning models. Lower-scale SPI
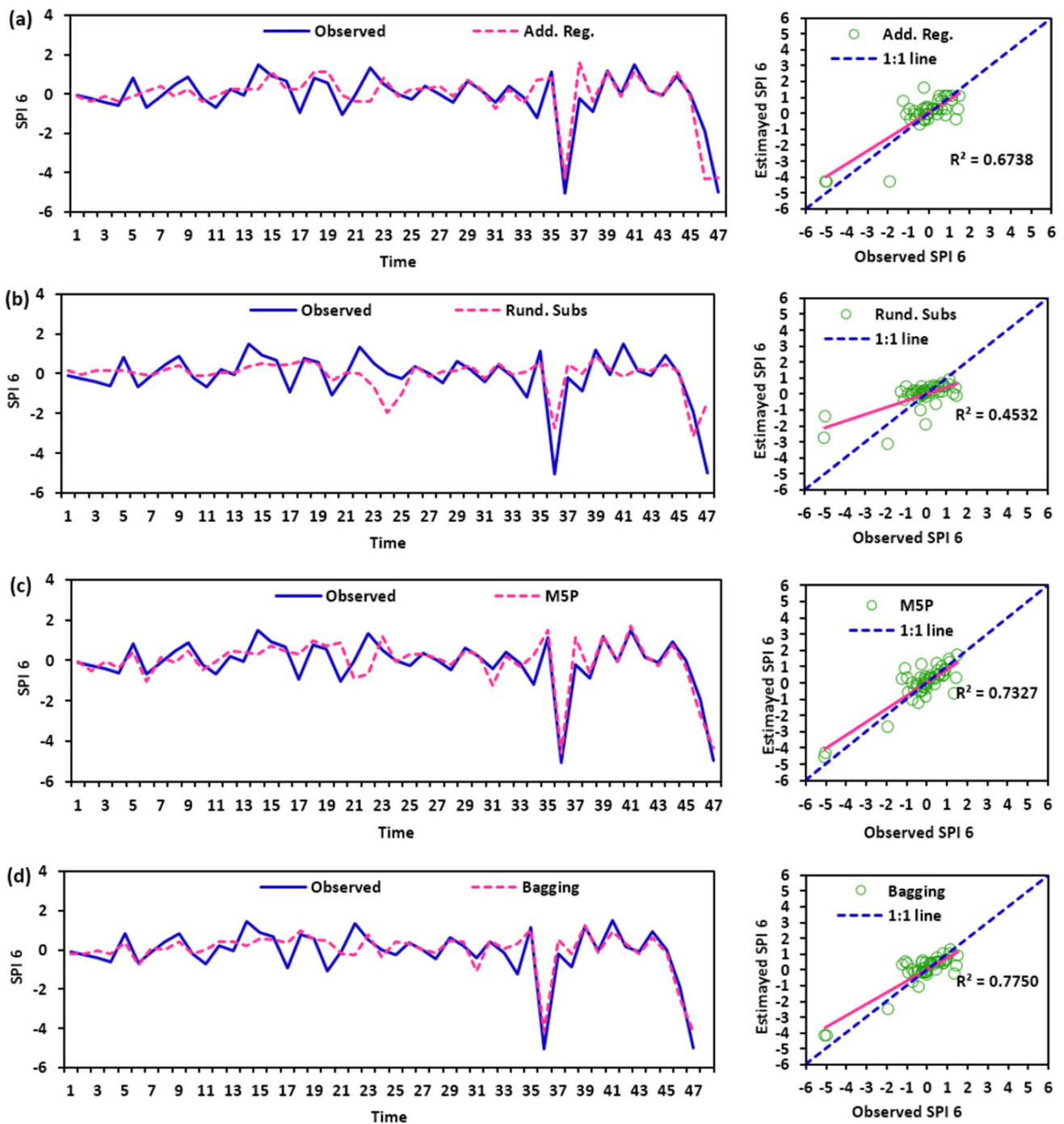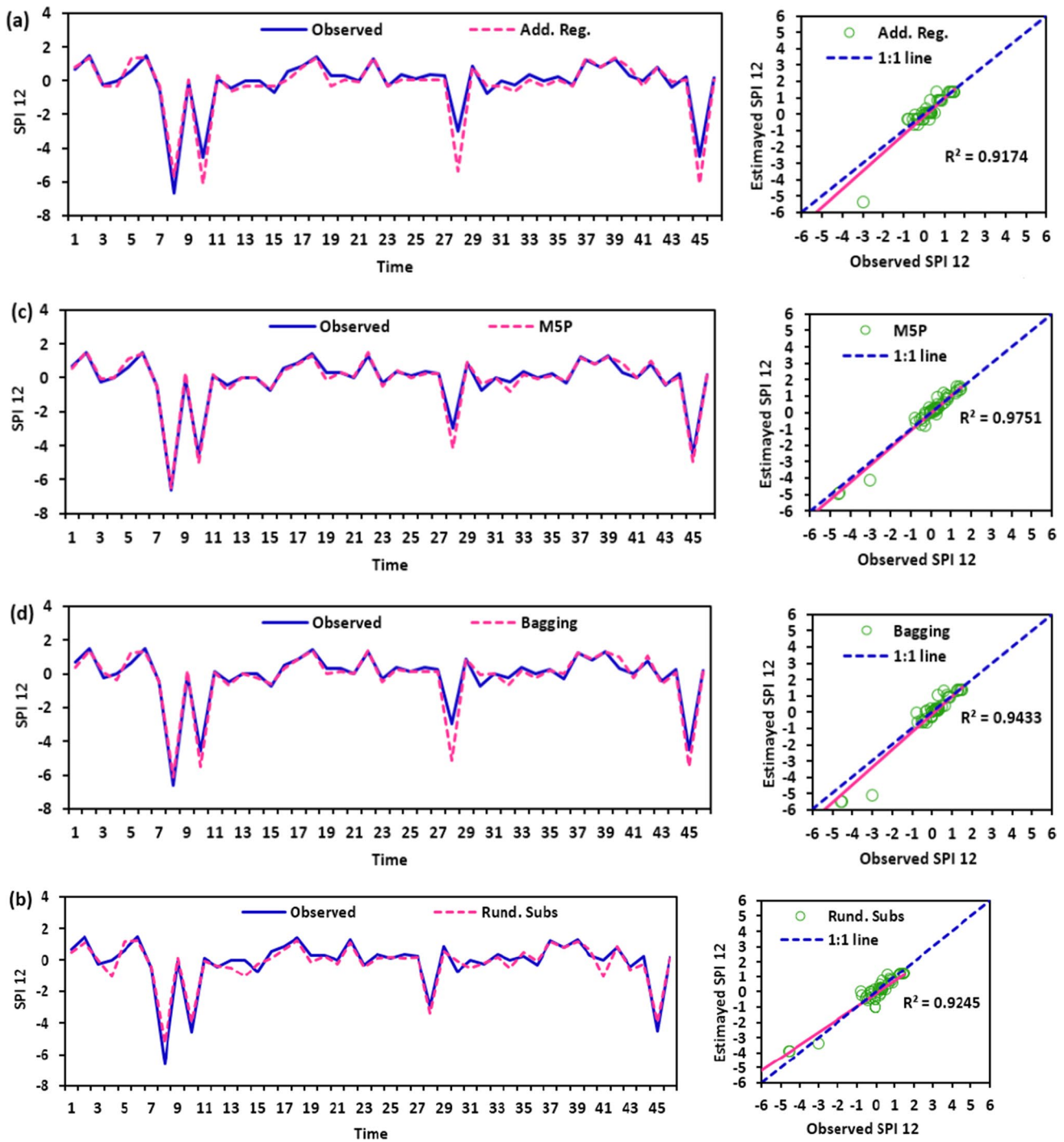
Fig. 10 Observed SPI versus Estimated SPI-3. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 3

predictions by M5P were nevertheless accurate. Using M5P and bagging, a very non-linear process may be recorded.

The M5P model is a multivariate linear algorithm. Numerous linear regression models are represented by the tree's leaves. This technique facilitates data segmentation and matching with the suitable regression model (Heddam and Kisi 2018). It is able to fit numerous models to diverse non-linear datasets because of its decomposition capacity. M5P was able to simulate every data point in a data series, which improved its capacity to foresee phenomena in linear models. This update significantly enhanced M5P's capacity to rapidly learn and model high-dimensional data.

Fig. 11 Observed SPI versus Estimated SPI-6. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 3

A stochastic (time-series) model and algorithms drawn from nature have been used to forecast a number of drought indicators. The outcomes of these models were contrasted with those of the M5P. (DIs). The forecast for the meteorological drought in Ankara, Turkey, was made using regression and random subspace models using delayed SPI data (Mehdizadeh et al. 2020). The prediction accuracy found in this investigation was higher than the predictive capability of the regression and random subspace models. Droughts in eastern Australia were predicted using the least-squares support vector machine (LSSVM), multivariate adaptive regression splines (MARS), and M5P tree models (Deo et al. 2017). The M5P tree technique was said to have better prediction accuracy. The ANFIS, M5P, M11, and M13 models were among the ML models employed by

Fig. 12 Observed SPI versus Estimated SPI-12. **a** Additive regression. **b** Random subspace. **c** M5P. **d** Bagging models during the testing period for station 3

Nguyen et al. (2015) to forecast SPI in the Cai River basin in Vietnam. The highest performing model, according to them, was M5P, which was followed by M11 and M13. Stepwise linear regression, genetic programming, and M5P approaches were used by Adarsh and Janga Reddy (2019) to forecast standardized precipitation indices for various areas of India. They observed that M5P performed better than expected in predicting droughts across the board. (Shamshirband et al. 2020) demonstrated improved results with the M5P and bagging models and predicted SPI using support vector regression, bagging, and M5P models. Barzkar et al. (2022) predicted SPIs for various climatic circumstances using three ML models: GEP, M5P, and
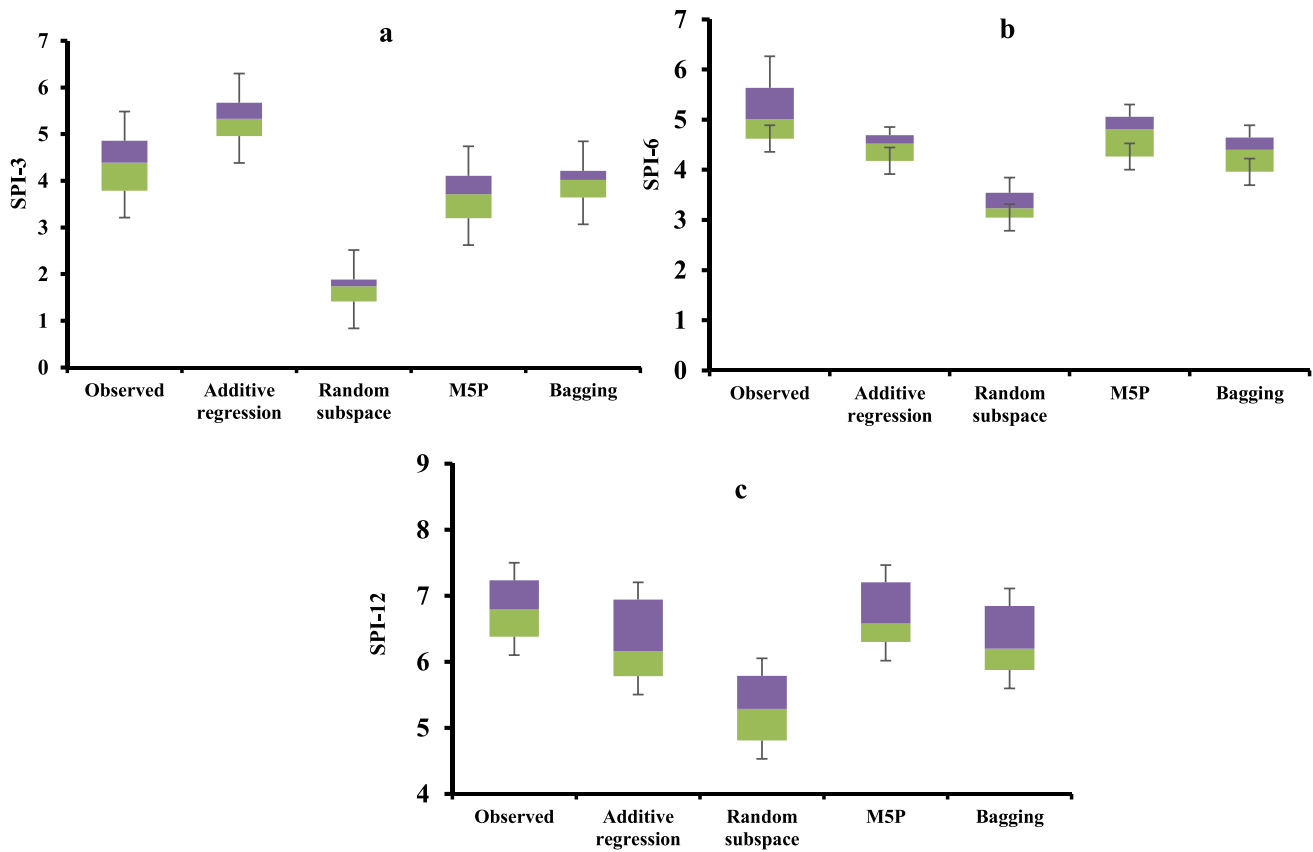
Fig. 13 Box plot presentation of the performance of the predictive models. **a** SPI-3. **b** SPI-6. **c** SPI-12 for station 3

multivariate adaptive regression spline (MARS). They demonstrated that the M5P model outperformed others in every instance (Elbeltagi et al. 2023b).

The material listed above unequivocally demonstrates the potential of ML models to forecast droughts in various meteorological contexts. According to the current study, the ML model-more particularly, M5P and bagging-was better able to predict meteorological droughts over a wide range of periods. The effects of droughts are disastrous to both society and the economy. The results of this study suggest that drought forecasting models might be placed as an alarm to lessen the consequences of drought in India's eastern areas, which are resistant to them.

5 Conclusions

In this study, four machine learning models namely an additive regression, random subspace, M5P, and bagging were selected to predict the future of SPI-3, SPI-6, and SPI-12 months at the Upper Godavari Basin, India. The input dataset series for the expansion of four models were pre-processed with machine learning to enhance the performance of the four models. Based on the statistical performance metrics, research results showed that the Bagging was the best model for predicting SPI-3 and SPI-6 while the M5P was the best for SPI-12 estimation in station 1, while in stations 12 and 13, the M5P was superlative in predicting the SPI-3 and SPI-12 months and the bagging was the best in SPI-6. The whole best models had acceptable forecasting of the mid-term drought forecasting based on the SPI-3, SPI-6, and SPI-12 months for three stations in the Upper Godavari Basin in India. Finally, these best machine learning models are better in predicting drought phenomena based on the standardized precipitation index (SPI) and it is not inadequate by the training input range and gives precise forecasts for short-term and mid-term drought situations. The results of the study area can be useful for making policy and planning related to drought, water resources management, crop water requirement, and irrigation planning purposes in the semi-arid region.

Author contribution Ahmed Elbeltagi and Chaitanya B. Pande had the original idea of the research. Chaitanya B. Pande: Conceptualization, Development of Methodology, Formal analysis, Original writing and

drafting, Writing—review and editing. Ahmed Elbeltagi: Conceptualization, Formal analysis, Software, Writing—review and editing. Romulus Costache, Saad Sh. Sammen, Rabeea Noor: Original writing and drafting, Writing—review and editing. All authors approved the final version for submission.

Declarations

Ethics approval The authors confirm that this article is original research and has not been published or presented previously in any journal or conference in any language (in whole or in part).

Consent to participate and consent to publish The authors declare that they have consent to participate and consent to publish.

Conflict of interest The authors declare no competing interests.

References

Adarsh S, Janga Reddy M (2019) Evaluation of trends and predictability of short-term droughts in three meteorological subdivisions of India using multivariate EMD-based hybrid modelling. Hydrol Process 33:130–143

Aghelpour P, Bahrami-Pichaghchi H, Kisi O (2020) Comparison of three different bio-inspired algorithms to improve ability of neuro fuzzy approach in prediction of agricultural drought, based on three different indexes. Comput Electron Agric 170:105279

Ali M, Deo RC, Downs NJ, Maraseni T (2018) An ensemble-ANFIS based uncertainty assessment model for forecasting multi-scalar standardized precipitation index. Atmospheric Res 207:155–180

Aragão LE, Anderson LO, Fonseca MG, Rosan TM, Vedovato LB, Wagner FH, Silva CV, Junior CHS, Arai E, Aguiar AP (2018) 21st century drought-related fires counteract the decline of Amazon deforestation carbon emissions. Nat Commun 9:1–12

Bahrami M, Bazrkar S, Zarei AR (2019) Modeling, prediction and trend assessment of drought in Iran using standardized precipitation index. J Water Clim Change 10:181–196

Barzkar A, Najafzadeh M, Homaei F (2022) Evaluation of drought events in various climatic conditions using data-driven models and a reliability-based probabilistic model. Nat Hazards 110:1931–1952

Belayneh A, Adamowski J (2012) Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. Appl Comput Intell Soft Comput Article ID 794061:13. https://doi.org/10.1155/2012/794061

Belayneh A, Adamowski J, Khalil B (2016) Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain Water Resour Manag 2:87–101. https://doi.org/10.1007/s40899-015-0040-5

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Carrão H, Russo S, Sepulcre-Canto G, Barbosa P (2016) An empirical standardized soil moisture index for agricultural drought assessment from remotely sensed data. Int J Appl Earth Obs Geoinformation 48:74–84

Choubin B, Malekian A, Golshan M (2016) Application of several data-driven techniques to predict a standardized precipitation index. Atmósfera 29:121–128

Dai F, Lee C, Ngai YY (2002) Landslide risk assessment and management: an overview. Eng Geol 64:65–87

Deo RC, Kisi O, Singh VP (2017) Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos Res 184:149–175. https://doi.org/10.1016/j.atmosres.2016.10.004

Dice J, Rodziewicz D (2020) Drought risk to the agriculture sector, federal reserve Bank of Kansas City, Econ Rev 105(2):61–86

Ditthakit P, Pinthong S, Salaeh N et al (2021) Using machine learning methods for supporting GR2M model in runoff estimation in an ungauged basin. Sci Rep 11:19955. https://doi.org/10.1038/s41598-021-99164-5

Domenikiotis C, Spiliotopoulos M, Tsiros E, Dalezios N (2004) Early cotton production assessment in Greece based on a combination of the drought Vegetation Condition Index (VCI) and the Bhalme and Mooley Drought Index (BMDI). Int J Remote Sens 25:5373–5388

Duan K, Sun G, Caldwell PV, McNulty SG, Zhang Y (2018) Implications of upstream flow availability for watershed surface water supply across the conterminous United States. JAWRA J Am Water Resour Assoc 54:694–707

Ebrahimpour M, Rahimi J, Nikkhah A, Bazrafshan J (2015) Monitoring agricultural drought using the standardized effective precipitation index. J Irrig Drain Eng 141:04014044

Elbeltagi A, Kumar M, Kushwaha NL et al (2023a) Drought indicator analysis and forecasting using data driven models: case study in Jaisalmer. India Stoch Environ Res Risk Assess 37:113–131. https://doi.org/10.1007/s00477-022-02277-0

Elbeltagi A, Pande CB, Kumar M et al (2023b) Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-023-25221-3

Finn JA, Suter M, Haughey E, Hofer D, Lüscher A (2018) Greater gains in annual yields from increased plant diversity than losses from experimental drought in two temperate grasslands. Agric Ecosyst Environ 258:149–153

Hänsel S, Schucknecht A, Matschullat J (2016) The Modified Rainfall Anomaly Index (mRAI)—is this an alternative to the Standardised Precipitation Index (SPI) in evaluating future extreme precipitation characteristics? Theor Appl Climatol 123:827–844

Heddam S, Kisi O (2018) Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J Hydrol 559:499–509

Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20:832–844

Hubbard KG, Wu H (2005) Modification of a crop-specific drought index for simulating corn yield in wet years. Agron J 97:1478–1484

Ibrahimi A, Baali A (2018) Application of several artificial intelligence models for forecasting meteorological drought using the standardized precipitation index in the Saiss Plain (Northern Morocco). Int J Intell Eng Syst 11:267–275

Jang SH, Lee J-K, Oh JH, Jo JW, Cho Y (2017) The probabilistic drought forecast based on the ensemble technique using the Korean surface water supply index. Nat Hazards Earth Syst Sci Discuss 1–51

Juhasz T, Kornfield J (1978) The Crop Moisture Index: unnatural response to changes in temperature. J Appl Meteorol 17:1864–1866

Khosravi K, Cooper JR, Daggupati P, Pham BT, Bui DT (2020) Bedload transport rate prediction: Application of novel hybrid data mining techniques. J Hydrol 585:124774

Komasi M, Sharghi S, Safavi HR (2018) Wavelet and cuckoo search-support vector machine conjugation for drought forecasting using

Standardized Precipitation Index (case study: Urmia Lake, Iran). J Hydroinformatics 20:975–988

Kumar Gautam V, Pande CB, Kothari M et al (2022) Exploration of groundwater potential zones mapping for hard rock region in the Jakham river basin using geospatial techniques and aquifer parameters. Adv Space Res. https://doi.org/10.1016/j.asr.2022.11.022

Liu C, Yang C, Yang Q, Wang J (2021) Spatiotemporal drought analysis by the standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI) in Sichuan Province. China Sci Rep 11:1–14

Lopez-Nicolas A, Pulido-Velazquez M, Macian-Sorribes H (2017) Economic risk assessment of drought impacts on irrigated agriculture. J Hydrol 550:580–589

Malik A, Tikhamarine Y, Sammen SS et al (2021a) Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. Environ Sci Pollut Res 28:39139–39158. https://doi.org/10.1007/s11356-021-13445-0

Malik A, Tikhamarine Y, Souag-Gamane D et al (2021b) Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. Meteorol Atmos Phys 133:891–909. https://doi.org/10.1007/s00703-021-00787-0

Mehdizadeh S, Ahmadi F, DanandehMehr A, Safari MJS (2020) Drought modeling using classic time series and hybrid wavelet-gene expression programming models. J Hydrol 587:125017

Meyer SJ, Hubbard KG, Wilhite DA (1993) A crop-specific drought index for corn: I Model development and validation. Agron J 85:388–395

Mohamadi S, Sammen SS, Panahi F et al (2020) Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. Nat Hazards 104:537–579. https://doi.org/10.1007/s11069-020-04180-9

Mokhtarzad M, Eskandari F, Vanjani NJ, Arabasadi A (2017) Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. Environ Earth Sci 76:1–10

Moron V (1994) Guinean and Sahelian rainfall anomaly indices at annual and monthly scales (1933–1990). Int J Climatol 14:325–341

Nguyen LB, Li QF, Ngoc TA, Hiramatsu K (2015) Adaptive neuro-fuzzy inference system for drought forecasting in the cai river basin in Vietnam. J Fac Agric Kyushu Univ 60:405–415

Ntale HK, Gan TY (2003) Drought indices and their application to East Africa. Int J Climatol J R Meteorol Soc 23:1335–1357

Orimoloye IR, Olusola AO, Belle JA et al (2022) Drought disaster monitoring and land use dynamics: identification of drought drivers using regression-based algorithms. Nat Hazards 112:1085–1106. https://doi.org/10.1007/s11069-022-05219-9

Pande CB, Al-Ansari N, Kushwaha NL, Srivastava A, Noor R, Kumar M, Moharir KN, Elbeltagi A (2022) Forecasting of SPI and meteorological drought based on the artificial neural network and M5P model tree land. 11(11):2040. https://doi.org/10.3390/land11112040

Pande CB, Kushwaha NL, Orimoloye IR et al (2023a) Comparative assessment of improved SVM method under different kernel functions for predicting multi-scale drought index. Water Resour Manage 37:1367–1399. https://doi.org/10.1007/s11269-023-03440-0

Pande CB, Kadam SA, Rajesh J, Gorantiwar SD, Shinde MG (2023b) Predication of sugarcane yield in the semi-arid region based on the sentinel-2 data using vegetation's indices and mathematical modeling. In: Pande CB, Moharir KN, Singh SK, Pham QB, Elbeltagi A (eds). Climate change impacts on natural resources, ecosystems and agricultural systems. Springer Climate. Springer, Cham. https://doi.org/10.1007/978-3-031-19059-9_12

Pande CB, Moharir KN (2023c) Application of hyperspectral remote sensing role in precision farming and sustainable agriculture under climate change: A review. In: Pande CB, Moharir KN, Singh SK, Pham QB, Elbeltagi A (eds). Climate change impacts on natural resources, ecosystems and agricultural systems. Springer Climate. Springer, Cham. https://doi.org/10.1007/978-3-031-19059-9_21

Pande CB, Moharir KN, Varade A (2023d) Water conservation structure as an unconventional method for improving sustainable use of irrigation water for soybean crop under rainfed climate condition. In: Pande CB, Moharir KN, Singh SK, Pham QB, Elbeltagi A (eds). Climate change impacts on natural resources, ecosystems and agricultural systems. springer climate. Springer, Cham. https://doi.org/10.1007/978-3-031-19059-9_28

Peng-cheng Q, Min L, Lan L (2016) Application of effective precipitation index in rainstorm flood disaster monitoring and assessment. Chin J Agrometeorol 37:84

Pham BT, Phong TV, Nguyen-Thoi T, Parial K, Singh SK, Ly H-B, Nguyen KT, Ho LS, Le HV, Prakash I (2020) Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. Geocarto Int 37(3):735–757. https://doi.org/10.1080/10106049.2020.1737972

Poornima S, Pushpalatha M (2019) Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. Soft Comput 23:8399–8412

Pramudya Y, Komariah, Dewi WS, Sumani, Mujiyo, Sukoco T A and Rozaki Z (2016) Remote sensing for estimating agricultural land use change as the impact of climate change (Proc of SPIE) 9877:987720–1.

Roodposhti MS, Safarrad T, Shahabi H (2017) Drought sensitivity mapping using two one-class support vector machine algorithms. Atmos Res 193:73–82

Sattari MT, Sureh FS (2019) Drought prediction based on standardized precipitation- evapotranspiration index by using M5 tree model. Int Civil Eng Archit Conf 1–14

Shamshirband S, Hashemi S, Salimi H, Samadianfard S, Asadi E, Shadkani S, Kargar K, Mosavi A, Nabipour N, Chau K-W (2020) Predicting standardized streamflow index for hydrological drought using machine learning models. Eng Appl Comput Fluid Mech 14:339–350

Shelar RS et al (2022) Sub-watershed prioritization of Koyna river basin in India using multi criteria analytical hierarchical process, remote sensing and GIS techniques. Phys Chem Earth 128:103219. https://doi.org/10.1016/j.pce.2022.103219

Soh Y, Koo C, Huang Y, Fung K (2018) Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin. Malaysia Comput Electron Agric 144:164–173

Sohrabi MM, Ryu JH, Abatzoglou J, Tracy J (2015) Development of soil moisture drought index to characterize droughts. J Hydrol Eng 20:04015025

Spennemann PC, Rivera JA, Saulo AC, Penalba OC (2015) A comparison of GLDAS soil moisture anomalies against standardized precipitation index and multisatellite estimations over South America. J Hydrometeorol 16:158–171

Stone CJ (1985) Additive regression and other nonparametric models. Ann Stat 13:689–705

Tan CP, Yang JP, Li M (2015) Temporal-spatial variation of drought indicated by SPI and SPEI in Ningxia Hui autonomous region. China Atmos 6(10):1399–1421

Tong S, Lai Q, Zhang J, Bao Y, Lusi A, Ma Q, Li X, Zhang F (2018) Spatiotemporal drought variability on the Mongolian Plateau from 1980–2014 based on the SPEI-PM, intensity analysis and Hurst exponent. Sci Total Environ 615:1557–1565

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J Climate 23:1696–1718

Wang X, Jiang D, Lang X (2017) Future extreme climate changes linked to global warming intensity. Sci Bull 62:1673–1680

Webber H, Ewert F, Olesen JE, Müller C, Fronzek S, Ruane AC, Bourgault M, Martre P, Ababaei B, Bindi M (2018) Diverging importance of drought stress for maize and winter wheat in Europe. Nat Commun 9:1–10

Wu H, Hayes MJ, Wilhite DA, Svoboda MD (2005) The effect of the length of record on the standardized precipitation index calculation. Int J Climatol J r Meteorol Soc 25:505–520

Xu B, Lin B (2015) Factors affecting carbon dioxide (CO_2) emissions in China's transport sector: a dynamic nonparametric additive regression model. J Clean Prod 101:311–322

Xu L, Abbaszadeh P, Moradkhani H, Chen N, Zhang X (2020) Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. Remote Sens Environ 250:112028

Yang Y, Zhang S, Roderick ML, McVicar TR, Yang D, Liu W, Li X (2020) Comparing Palmer Drought Severity Index drought assessments using the traditional offline approach with direct climate model outputs. Hydrol Earth Syst Sci 24:2921–2930

Yariyan P, Janizadeh S, Phong TV, Nguyen HD, Costache R, Le HV, Pham BT, Pradhan B, Tiefenbacher JP (2020) Improvement of best first decision trees using bagging and dagging ensembles for flood-risk mapping. Water Resour Manag. https://doi.org/10.1007/s11269-020-02603-7

Yaseen ZM, Ali M, Sharafati A, Al-Ansari N, Shahid S (2021) Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh. Sci Rep 11:1–25

Yu H, Zhang Q, Xu C-Y, Du J, Sun P, Hu P (2019) Modified palmer drought severity index: model improvement and application. Environ Int 130:104951

Zhan C, Gan A, Hadi M (2011) Prediction of lane clearance time of freeway incidents using the M5P tree algorithm. IEEE Trans Intell Transp Syst 12:1549–1557