



Drought indicator analysis and forecasting using data driven models: case study in Jaisalmer, India

Ahmed Elbeltagi¹ · Manish Kumar² · N. L. Kushwaha³ · Chaitanya B. Pande⁴ · Pakorn Ditthakit⁵ · Dinesh Kumar Vishwakarma⁶ · A. Subeesh⁷

Accepted: 4 July 2022 / Published online: 1 August 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Agricultural droughts are a prime concern for economies worldwide as they negatively impact the productivity of rain-fed crops, employment, and income per capita. In this study, Standard Precipitation Index (SPI) has been used to evaluate different drought indices for Rajasthan of India. In agricultural, hydrological, and meteorological applications such as irrigation scheduling, crop simulation, water budgeting, reservoir operations, and weather forecasting, the accurate estimation of the drought indices such as the Standardized Precipitation Index (SPI) plays an important role. Thus, the present study was conducted to examine the feasibility and effectiveness of the Random Subspace (RSS) model and its hybridization with the M5 Pruning tree (M5P), Random Forest (RF), and Random Tree (RT) to estimate the SPI at 3, 6, and 12 droughts during 2000–2019. Performances of RSS and hybridized algorithms were assessed and compared using performance indicators (i.e., MAE, RMSE, RAE, RRSE, and R^2) and various graphical interpretations. Results indicated that the RSS-M5P provided the most accurate SPI prediction (MAE = 0.497, RMSE = 0.682, RAE = 81.88, RRSE = 87.22, and $R^2 = 0.507$ for SPI-3; MAE = 0.452, RMSE = 0.717, RAE = 69.76, RRSE = 85.24, and $R^2 = 0.402$ for SPI-6. And MAE = 0.294, RMSE = 0.377, RAE = 55.79, RRSE = 59.57, and $R^2 = 0.783$ for SPI-12) compare to RSS alone, RSS-RF, and RSS-RT models for study the drought situation in Jaisalmer Rajasthan. The M5P algorithms have improved the performance of the RSS structure.

Keywords SPI · Random Subspace · Random Tree · Subset regression · Sensitivity analysis

Abbreviations

SPI	Standard Precipitation Index
RSS	Random Subspace
M5P	M5 Pruning tree
RF	Random Forest
RT	Random Tree
MAE	Mean absolute error
RMSE	Root mean square error
RAE	Relative absolute error
RRSE	Root relative squared error
R^2	Coefficient of determination
P	Precipitation
PET	Potential evapotranspiration
°C	Degree celsius
mm	Millimetre
DrinC	Drought Indices Calculator
SDR	Standard deviation reduction factor
SE	Standard error

ML	Machine learning
MPMR	Minimum probability machine regression
ELM	Extreme learning machine
OSELM	Online sequential-ELM
ANNs	Artificial neural networks
SVR	Support vector regression
WA-ANN	Coupled wavelet-anns
MLP	Multilayer perceptron
MLP-ICA	Imperialistic Competitive Algorithm-MLP
MSPI	Multivariate Standardized Precipitation Index

1 Introduction

Drought is a natural occurrence that affects various climatic, hydrological, and environmental systems, which all have socioeconomic impacts (Vicente-Serrano et al. 2020). Drought is generally caused by meteorological irregularities, such as lower rainfall spells that result in water

shortages at particular points of the water cycle or throughout the cycle (McKee et al. 1993). The world's increasing demand for scarce water resources has been rising rapidly by challenging its availability for food production and other vital purposes while putting sustainable development at risk (Kushwaha et al. 2016; Abd-Elaty et al. 2022a). As one of the primary water consumers upon which a growing population depends, agriculture competes with household, industrial, and environmental uses for these scarce water resources (Suwal et al. 2020; Kuriqi et al. 2020b; Kushwaha et al. 2022a). Terrestrial evapotranspiration is essential to a wide range of the atmosphere, hydrosphere, and biosphere processes to connect with the water, energy, and carbon cycles. These factors play an important role in the hydrologic cycle, flora and fauna community dynamics, and many biochemical processes (Kuriqi et al. 2020a; Abd-Elaty et al. 2022b). Several aspects of sustainable water resources management depend upon the precise estimation of terrestrial evapotranspiration. Even though natural environmental change versus anthropogenic activities differs among regions (Anand et al. 2018; Elbeltagi et al. 2022a), the role of climate change impacts on the hydrologic cycle has been characterized and recognized mostly across different global regions (Kushwaha et al. 2022b). It is widely accepted among scientific communities and policymakers that climate change significantly influences water availability, especially in arid and semi-arid regions (Mao et al. 2015; Elbeltagi et al. 2020, 2021). Many models have shown that terrestrial evapotranspiration varies in magnitudes at both temporal and spatial scales (Chen et al. 1996; Wang and Dickinson 2012). In general, all satellite-based products demonstrate a significant increase in terrestrial evapotranspiration over the past three decades (Rodell et al. 2009; Zeng et al. 2018). The empirical-based methods such as the Penman-Monteith equation are the most commonly applied in practice mainly due to their application simplicity.

The empirical-based approaches in the same line as the one mentioned above are, however, inaccurate only when the vegetation is not water-stressed and the net radiation and stomatal resistance are available (Tao et al. 2018; Elbeltagi et al. 2022b). Various other approaches make use of recent satellite data. Indeed, among many terrestrial evapotranspiration estimation methods, satellite remote sensing-based data demonstrate to be one of the most promising ways of terrestrial evapotranspiration mapping over larger areas (Yang et al. 2013). However, both approaches, i.e., empirical and satellite-based methods, fail to monitor long-term global terrestrial evapotranspiration (Wang et al. 2010; Anand et al. 2018). Therefore, robust estimation of the long-term variability of terrestrial evapotranspiration requires the use of standard meteorological data supplemented with high-resolution satellite data (Chen

et al. 1996; Wang et al. 2010). Despite the improvement of meteorological observation's spatial resolution, it is worth exploring methods to estimate terrestrial evapotranspiration not only from remotely sensed information but also considering data-driven based techniques combined with remote sensing data (Yang et al. 2013). This approach would considerably condense the input effort and increase the robustness of satellite-based terrestrial evapotranspiration models while helping policymakers establish sustainable water resources management strategies. Several studies estimated evapotranspiration applying data-driven techniques (Chen et al. 1996; Muhammad Adnan et al. 2020; Alizamir et al. 2020). Nevertheless, to the best of the author's knowledge, there is a lack of studies that combine the application of data-driven techniques with satellite-based observation to estimate terrestrial evapotranspiration over complex river basins similar to the one considered in this study. India has a large number of underprivileged drought-prone and economically backward regions, but no effective system of monitoring and assessing drought for the development of better management strategies and policies, and such studies overstate meteorological trends spatially, temporally, and over time. An effective adaptation and mitigation strategy to overcome a drought situation in a country like India will depend on the identification of a reliable methodology for calculating the drought index.

SPI may be calculated over any period between one month and 72 months. A practical range of application of 1–24 months is best based on statistical evidence (Poornima and Pushpalatha 2019). This 24-month cutoff is based on Guttman's recommendation of having around 50–60 years of data available. It is impossible to do statistical analysis on the tails (both wet and dry extremes) without having more than 80–100 years of data. A monthly SPI can be calculated in theory, but doing so in practice is not recommended. An average period of at least four weeks is highly recommended for the user. Even in non-arid climates, one is likely to encounter many dry days (0.00 rainfall even in non-arid climates) that cause the SPI to behave very erratically (Thomas et al. 2015); therefore, this approach is not recommended. Nonetheless, updating the SPI daily or weekly for a 1-month up to a 24-month period is acceptable. Accordingly, we used 3, 6 and 12 SPI Index in this study to achieve better results.

In light of the above facts, to the author's knowledge, no study has been carried out to utilize the capability of the Hybrid Random SubSpace model with other machine learning algorithms for predicting extreme droughts in the studied area in India. The primary objective of this work was to develop a new model for predicting extreme droughts based on the SPI. SPI is used in this study because it combines meteorological variables [i.e., precipitation

(P) and potential evapotranspiration (PET)] and shows drought better than any other index (Ali et al. 2019).

The present study was conducted to examine the feasibility and effectiveness of the Random Subspace (RSS) to estimate the SPI 3, 6 and 12 months in India, during 2000–2019. In this study, several inputs were constructed and the best subset regression was applied to select the most effective variables as inputs to the developed artificial models. The present study aims to create hybrid data-driven models coupled with RSS using the stacking hybridization technique and evaluate their performance. Consequently, the proposed drought prediction models can help in agricultural, hydrological, and meteorological applications such as irrigation scheduling, crop simulation, water budgeting, reservoir operations, and weather forecasting. The remaining sections of the manuscript are organized as follows: Sect. 2, presents a short introductory information about the study site, data collection and methodology; Sect. 3, summarizes the main findings of the study; Sect. 4, discusses the practical implications of the main findings; Sect. 5, summarizes the main conclusions of the study. The main objectives of this study area are as follows: (1) To develop the machine learning models based on the 20 years' datasets of rain gauge. (2) To performance of ML algorithms in arid climatic condition for prediction of drought monitoring. (3) Comparison of ML models, and recommendations for prediction of drought.

2 Materials and methods

2.1 Study area and data acquisition

The study site Jaisalmer is the largest district of Rajasthan (India) situated at a latitude of 26°55' N and longitude of 70°55' E (Fig. 1). The Jaisalmer is also nicknamed “The Golden City”. The study site is situated at the heart of the “Thar Desert” (the Great Indian Desert). It comes under an arid region and is prone to temperature extremes. The temperature varies greatly from day to night in both summer and winter. The maximum summer temperature is around 49 °C while the minimum is 25 °C. While temperature during winter varies from 23.6 to 5 °C. The average rainfall was observed to be scanty i.e., 201 mm with average rainy days as 12 in a year. In the present study, 20 years (2000–2019) of rainfall data were used for SPI evaluation. Further, SPI estimation models used 70 per cent data as training and 30 per cent data as testing the models.

2.2 Methodology

2.2.1 SPI description and calculation

Various indices may be used to analyses and monitor droughts across a large area, each with its own set of strengths and disadvantages. SPI, the most generally used and acknowledged index, is based on the premise that a decrease in precipitation relative to normal precipitation is the major predictor of droughts (McKee et al. 1993). The SPI is usually calculated monthly, such as 1, 3, 6, 9, and 12 months, and the drought intensity is determined using the SPI data. 6, 9 and 12 months, and the drought intensity is defined based on the calculated SPI values.

The probability density function of the total precipitation is used to calculate SPI. For every month and every location, this is done differently. Then, the probability functions have been transformed into the standard normal distribution. The probability function was used to express the gamma distribution:

$$gg(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} X^{\alpha-1} e^{-x/\beta} \quad (1)$$

Here α denotes the appearance of parameters, β represents the range of the parameter, x denotes the amount of rainfall, and $\Gamma(\alpha)$ is the gamma function. The value of α and β parameters > 0 . The Gamma function $\Gamma(\alpha)$ can be expressed as follows:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \quad (2)$$

α and β parameters must be estimated for adjusting the gamma distribution. Maximum likelihood solutions are used to accurately obtain α and β as follows:

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

$$\hat{\beta} = \frac{\bar{X}}{\hat{\alpha}} \quad (4)$$

where

$$A = \ln(\bar{X}) - \frac{\sum \ln(\bar{X})}{n} \quad (5)$$

The SPI has strengths such as It is flexible: it can be computed for multiple timescales; Shorter timescale SPIs, for example 1-, 2- or 3-month SPIs, can provide early warning of drought and help assess drought severity; It is spatially consistent: it allows for comparisons between different locations in different climates and Its probabilistic nature gives it historical context, which is well suited for decision-making. There are few limitations like it is based only on precipitation and no soil water-balance component,

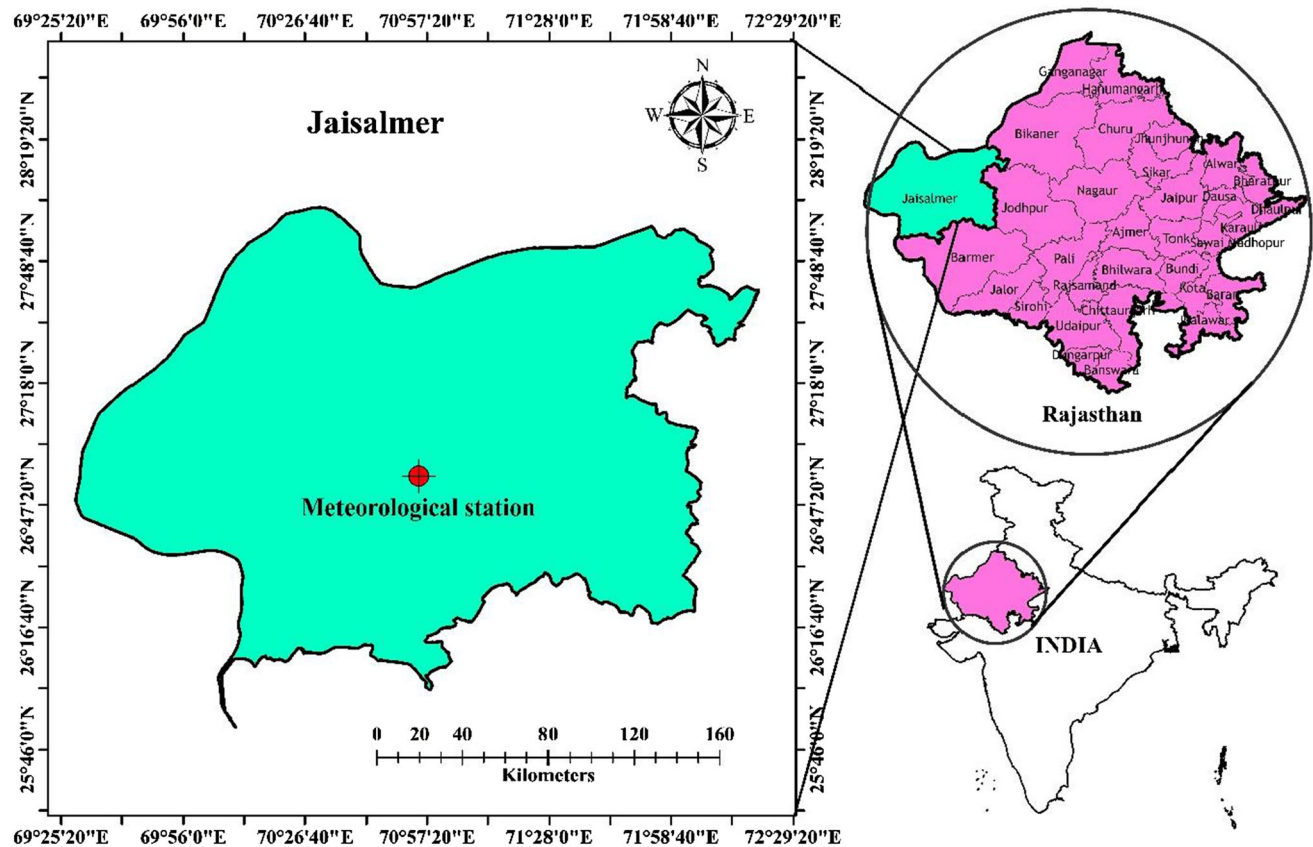


Fig. 1 Location map of the study station

thus no ratios of evapotranspiration/potential evapotranspiration (ET/PET) can be calculated.

The present study applied DrinC (Drought Indices Calculator) for the calculation of SPI (3, 6 and 12) and the calculated SPI used as a reference value for the performance assessment of the hybrid metaheuristics algorithms i.e., Random subspace (RSS), M5P, Random Forest (RF) and Random Tree (RT). The DrinC model was developed at the Laboratory of National Technical University, Athens (Tigkas et al. 2015).

2.2.2 Machine learning models

2.2.2.1 Random subspace

The random subspace method (RSM) is an ensemble learning technique presented by Ho (1998) to enhance the performance of the weak learners and improve the accuracy of individual learners (Pham et al. 2018). Random subspace builds a set of feature subspaces using random sampling and then trains the basic classifiers on top of them. The models are trained parallel, and multiple results are generated before being aggregated into the final results (Dong et al. 2020). The most important parameters used in this model include the number of seeds and the number of iterations. The optimal values of these

parameters are often achieved based on trial and error (Mosavi et al. 2020). Modification to the training data is performed in RSM, however this modification is applied in the feature space. Once there data consists of multiple redundant features, an improved model can be found in the random subspaces compared to the original feature space. The random subspace method is found to be the best performing when there are a large number of features, and discriminative information is spread across them. On the other hand, when there are less informative features and data is noisy, the random subspace method tends to underperform (Sammut and Webb 2017). Figure 2 shows the block diagram of the random subspace model.

2.2.2.2 M5P

M5P algorithm resulted from the reconstruction of Quilan's M5 algorithm (1992). In M5P, linear regression functions are incorporated into the leaf nodes along with the functionality of a traditional decision tree. To construct the tree, the M5P algorithm utilizes the decision tree induction algorithm, and the splitting condition minimizes the amounts of intra-subset variation present on each branch of the tree. There are four main steps in the M5P algorithm. M5P splits the input space into multiple subspaces at the beginning of the process. These

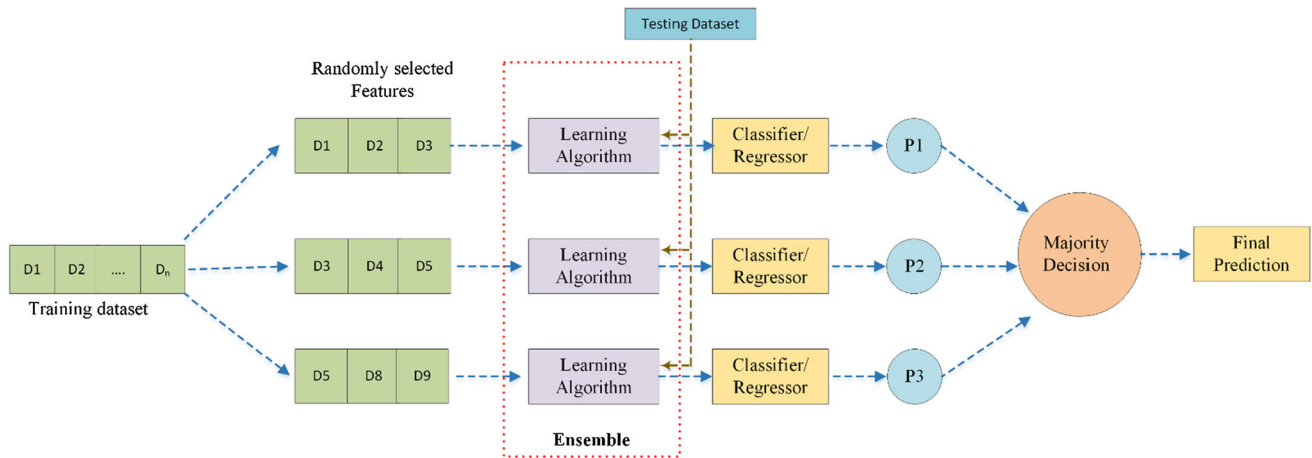


Fig. 2 Block diagram of random subspace model

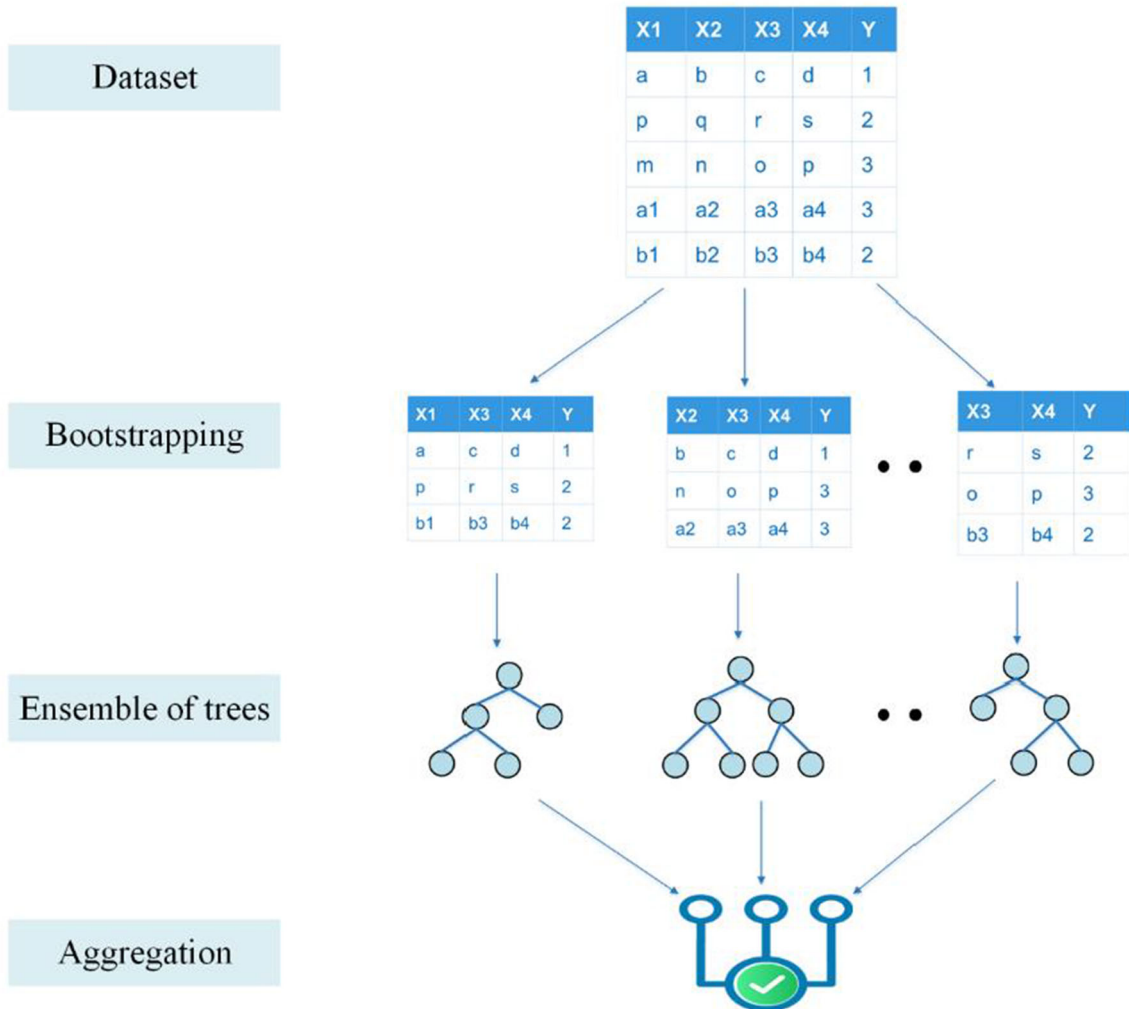


Fig. 3 Steps involved in solving a classification/regression problem using random forest

subspaces are used for building the tree. Using the standard deviation reduction factor (SDR), we construct the tree by minimizing the intra-subspace variability and intra-

subspace variability with a splitting criterion. As a result, the SDR factor will maximize the expected error reduction at the node (Wang and Witten 1996). In the second stage, a

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

No. of variables	Variables	MS E	R ²	Adjusted R ²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
(A) SPI-3								
1	SPI-1	0.507	0.413	0.410	5.517	- 151.307	- 144.466	0.592
2	SPI-1/SPI-3	0.492	0.433	0.428	- 0.201	- 157.107	- 146.846	0.577
3	SPI-1/SPI-3/SPI-4	0.490	0.438	0.430	- 0.100	- 157.068	- 143.386	0.577
4	SPI-1/SPI-3/SPI-4/SPI-9	0.491	0.440	0.430	1.072	- 155.928	- 138.825	0.580
5	SPI-1/SPI-3/SPI-4/SPI-8/SPI-9	0.490	0.443	0.430	1.885	- 155.168	- 134.645	0.582
6	SPI-1/SPI-3/SPI-4/SPI-6/SPI-8/SPI-9	0.492	0.444	0.429	3.538	- 153.531	- 129.587	0.586
7	SPI-1/SPI-3/SPI-4/SPI-5/SPI-6/SPI-8/SPI-9	0.493	0.445	0.427	5.148	- 151.940	- 124.576	0.591
8	SPI-1/SPI-3/SPI-4/SPI-5/SPI-6/SPI-8/SPI-9/SPI-10	0.495	0.445	0.425	7.057	- 150.036	- 119.251	0.596
9	SPI-1/SPI-2/SPI-3/SPI-4/SPI-5/SPI-6/SPI-8/SPI-9/SPI-10	0.497	0.445	0.422	9.028	- 148.067	- 113.861	0.601

Table 1 continued

10	SPI-1/SPI-2/SPI-3/SPI-4/SPI-5/SPI-6/SPI-7/SPI-8/SPI-9/SPI-10	0.500	0.445	0.419	11.000	- 146.096	- 108.470	0.606
(B) SPI-6								
1	SPI-1	0.412	0.583	0.581	14.666	- 195.529	- 188.715	0.420
2	SPI-1/SPI-6	0.400	0.598	0.594	8.671	- 201.294	- 191.072	0.410
3	SPI-1/SPI-3/SPI-6	0.391	0.609	0.603	4.381	- 205.597	- 191.969	0.402
4	SPI-1/SPI-3/SPI-6/SPI-7	0.388	0.614	0.607	3.600	- 206.442	- 189.406	0.400
5	SPI-1/SPI-3/SPI-6/SPI-7/SPI-9	0.384	0.619	0.610	2.557	- 207.598	- 187.155	0.398
6	SPI-1/SPI-3/SPI-5/SPI-6/SPI-7/SPI-9	0.384	0.621	0.610	3.721	- 206.472	- 182.622	0.400
7	SPI-1/SPI-2/SPI-3/SPI-5/SPI-6/SPI-7/SPI-9	0.385	0.622	0.609	5.201	- 205.017	- 177.760	0.403
8	SPI-1/SPI-2/SPI-3/SPI-4/SPI-5/SPI-6/SPI-7/SPI-9	0.386	0.622	0.608	7.013	- 203.215	- 172.551	0.406
9	SPI-1/SPI-2/SPI-	0.388	0.622	0.606	9.003	- 201.226	- 167.155	0.410

Table 1 continued

	3/SPI- 4/SPI- 5/SPI- 6/SPI- 7/SPI- 8/SPI-9							
10	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI- 5/SPI- 6/SPI- 7/SPI- 8/SPI- 9/SPI-10	0.39 0	0.62 2	0.604	11.000	- 199.22 9	- 161.750	0.413
(C) SPI-12								
1	SPI-1	0.16 6	0.83 6	0.835	21.917	- 387.93 1	- 381.171	0.166
2	SPI- 1/SPI-10	0.15 1	0.85 1	0.850	2.113	- 407.06 9	- 396.929	0.152
3	SPI- 1/SPI- 2/SPI-10	0.14 8	0.85 5	0.853	- 1.745	- 411.11 7	- 397.597	0.149
4	SPI- 1/SPI- 2/SPI- 3/SPI-10	0.14 8	0.85 6	0.853	- 0.668	- 410.08 6	- 393.186	0.150
5	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI-10	0.14 8	0.85 6	0.853	1.107	- 408.32 2	- 388.043	0.151
6	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI- 7/SPI-10	0.14 9	0.85 6	0.852	3.027	- 406.40 7	- 382.747	0.152
7	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI- 6/SPI- 7/SPI-10	0.15 0	0.85 6	0.851	5.014	- 404.42 0	- 377.381	0.153
8	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI-	0.15 0	0.85 6	0.851	7.011	- 402.42 3	- 372.004	0.155

Table 1 continued

	6/SPI- 7/SPI- 8/SPI-10							
9	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI- 6/SPI- 7/SPI- 8/SPI- 9/SPI-10	0.15 1	0.85 6	0.850	9.000	- 400.43 5	- 366.636	0.156
10	SPI- 1/SPI- 2/SPI- 3/SPI- 4/SPI- 5/SPI- 6/SPI- 7/SPI- 8/SPI- 9/SPI-10	0.15 2	0.85 6	0.849	11.000	- 398.43 5	- 361.256	0.158

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

linear regression model is developed for each subspace using the sub-dataset. In the third stage, the developed tree’s nodes are pruned to avoid overfitting problems, and the discontinuities induced by the pruning process are removed using a smoothing procedure in the final stage (Melesse et al. 2020). The key advantage of using M5P is that they are efficient in handling large datasets with higher dimensions. M5P was found to be robust while dealing with missing data. (Behnood et al. 2017).

2.2.2.3 Random Forest Random forest (RF) is an ensemble method that uses several decision trees parallel with the bagging (bootstrapping followed by aggregation) approach. RF was introduced by Breiman (2001) and used in numerous studies. Bootstrapping indicates that various individual decision trees are trained in parallel on different subsets of the input training dataset. This reduces the overall variance of the model and produces accurate results. For making the final decision in RF, the decisions of individual trees are aggregated, ensuring better generalization (Misra and Li 2020). In general, deep decision trees suffer from overfitting, and random forest prevents this by generating random subsets of features and building

smaller trees using those subsets. A random forest’s generalization error is based on the strength of the individual trees built and their correlations. Several studies have demonstrated that random forest models can effectively predict and estimate small sample sizes and complex data in both classification and regression problems (Biau and Scornet 2015). One of the key characteristics of random forest is resilience towards overfitting. As there are enough trees in the random forest model, the model gives better generalization. One major downside of random forest is that when the number of trees is very high, the algorithm gets slower. Steps involved in solving a classification/regression problem using the random forest are shown in Fig. 3.

2.2.2.4 Random Tree Machine learning is primarily based on random trees (RT), which combine single model trees and random forests. A random set of data is produced using a bagging approach through RT to have many individual learners and handle regression and classification problems (Kalmegh 2015). Since RT combines both bagging and RF methods, it generates a final prediction based on the aggregated prediction of multiple individual trees.

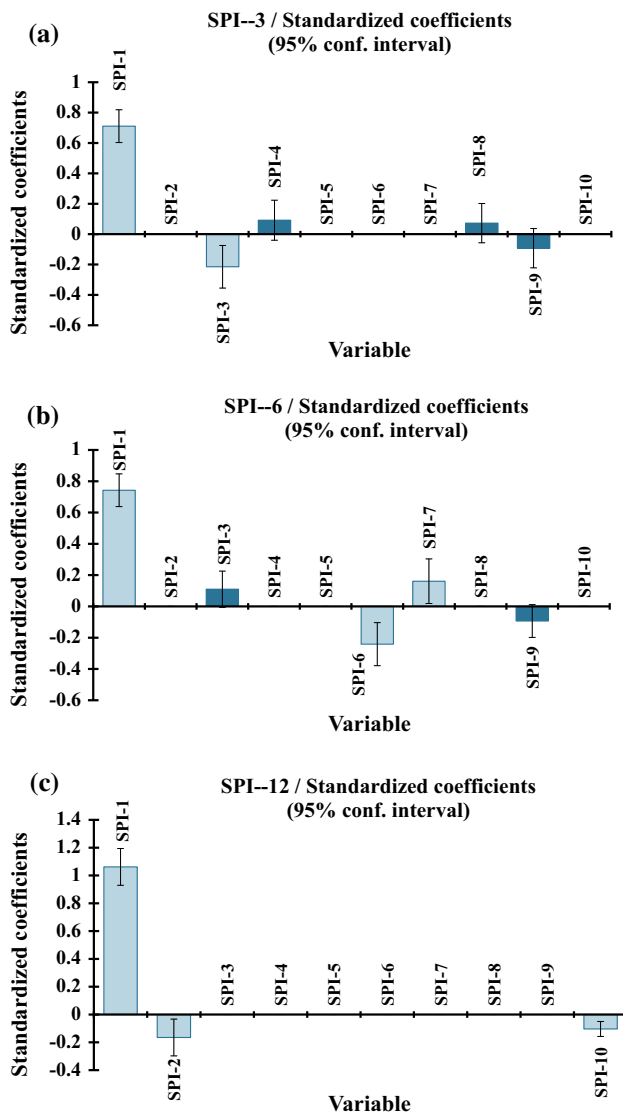


Fig. 4 The standardized coefficients of input variable for sensitivity analysis selected meteorological station for **a** SPI-3, **b** SPI-6, and **c** SPI-12

The RT algorithm starts by splitting the dataset into subspaces and fitting a constant to each subspace. In the first step, random tree classification algorithm takes the input feature vector, and classification is performed at each tree in the forest. A single tree model tends to perform poorly but bagging RT shows high performance in terms of accuracy. RTs have increased flexibility and enhanced training capability (Nhu et al. 2020) in a number of real-world problems. Additionally, RT maintains accuracy even after changing the model complexity. The single leaves on the trees indicate the linear models can be optimized via replacement methods or by splitting the attributes at every node to identify the best means to split a subset during optimization (Khosravi et al. 2019).

2.3 Hybridization of machine learning algorithms using stacked generalization

The SPI was predicted using stacking hybrid algorithms in this study. Wolpert (1992) proposed a stacking hybrid algorithm technique. During the training period, this method provides an environment for ensemble algorithms, which mix two or more algorithms. According to studies (Healey et al. 2018; Rahman et al. 2021) stacking hybrid algorithms can improve algorithm predictability. The idea behind stacking hybrid generalization is to use first-level learners to train and forecast training data sets. The first level learners' projected results were combined to create a new training dataset for the meta learner. Sikora et al. (2015) and Zhou (2009) provided more details on stacked hybrid generalization.

2.4 Best subset regression and sensitivity analysis

2.4.1 Input selection using best subset model for the SPI 3, 6, and 12 months of a selected station

Best subsets regression is an exploratory model building regression analysis. It compares all possible models that can be created based upon an identified set of predictors. It aims to find a small subset of predictors, so that the resulting linear model is expected to have the most desirable prediction accuracy several statistical criteria have been used to select the best combination of inputs, i.e., MSE, determination coefficients (R^2), adjusted R^2 , Mallows' Cp, Akaike's AIC, Schwarz's SBC, and Amemiya's PC. One of the most crucial procedures for developing the predictive model of multi-step ahead SPI drought index is to select the best subset combination of input data. Based on ten inputs, Table 1 shows (a) SPI-3, (b) SPI-6, and (c) SPI-12 for predicting the lag-time SPI. The seven statistical criteria have been used to select the best combination of inputs, i.e., MSE, determination coefficients (R^2), adjusted R^2 , Mallows' Cp, Akaike's AIC, Schwarz's SBC, and Amemiya's PC. The best subset input combination is displayed in bold blue row due to providing the lowest values of MSE, Mallows' Cp, Akaike's AIC, Schwarz's SBC, and Amemiya's PC, and the highest values of R^2 and Adjusted R^2 . For SPI-3, the best input combination is of 5 variables, i.e., SPI-1, SPI-3, SPI-4, SPI-8, and SPI-9. It provides MSE of 0.490, R^2 of 0.443, Adjusted R^2 of 0.430, Mallows' Cp of 1.885, Akaike's AIC of -155.168 , Schwarz's SBC of 134.645, and Amemiya's PC of 0.582. For SPI-6, the best input combination is of 5 variables, i.e., SPI-1, SPI-3, SPI-6, SPI-7, and SPI-9. It gives MSE of 0.384, R^2 of 0.619, Adjusted R^2 of 0.610, Mallows' Cp of

Table 2 The regression analysis for identifying the most effective parameters at Paithan station

Source	Value	Standard error	t	Pr > t	Lower bound (95%)	Upper bound (95%)
(A) SPI-3						
SPI-1	0.711	0.055	12.987	< 0.0001	0.603	0.819
SPI-2	0.000	0.000				
SPI-3	− 0.215	0.071	− 3.021	0.003	− 0.355	− 0.075
SPI-4	0.092	0.067	1.372	0.171	− 0.040	0.223
SPI-5	0.000	0.000				
SPI-6	0.000	0.000				
SPI-7	0.000	0.000				
SPI-8	0.072	0.066	1.100	0.273	− 0.057	0.202
SPI-9	− 0.093	0.066	− 1.411	0.160	− 0.222	0.037
SPI-10	0.000	0.000				
(B) SPI-6						
SPI-1	0.742	0.053	13.996	< 0.0001	0.638	0.847
SPI-2	0.000	0.000				
SPI-3	0.110	0.059	1.870	0.063	− 0.006	0.226
SPI-4	0.000	0.000				
SPI-5	0.000	0.000				
SPI-6	− 0.242	0.070	− 3.458	0.001	− 0.379	− 0.104
SPI-7	0.161	0.072	2.226	0.027	0.018	0.303
SPI-8	0.000	0.000				
SPI-9	− 0.093	0.053	− 1.759	0.080	− 0.198	0.011
SPI-10	0.000	0.000				
(C) SPI-12						
SPI-1	1.062	0.067	15.825	< 0.0001	0.929	1.194
SPI-2	− 0.165	0.067	− 2.453	0.015	− 0.298	− 0.032
SPI-3	0.000	0.000				
SPI-4	0.000	0.000				
SPI-5	0.000	0.000				
SPI-6	0.000	0.000				
SPI-7	0.000	0.000				
SPI-8	0.000	0.000				
SPI-9	0.000	0.000				
SPI-10	− 0.104	0.027	− 3.808	0.000	− 0.158	− 0.050

2.557, Akaike’s AIC of − 207.598, Schwarz’s SBC of − 187.155, and Amemiya’s PC of 0.398. For SPI-12, the best input combination is of 3 variables, i.e., SPI-1, SPI-2, and SPI-10. It has MSE of 0.148, R² of 0.855, Adjusted R² of 0.853, Mallows’ Cp of − 1.745, Akaike’s AIC of − 411.117, Schwarz’s SBC of − 397.597, and Amemiya’s PC of 0.149.

2.4.2 Sensitivity analysis

The combinations of the input variables strongly influence the models’ performance. Some contribute positively to the accuracy of the selected model, while others may contribute negatively. Sensitivity analysis was used to choose the most influential variables to optimize model

performance in predicting the SPI drought index. A regression analysis was performed at Jaisalmer to identify the most effective parameter sets. As shown in Fig. 4, the standardized coefficients of input variables for (a) SPI-3, (b) SPI-6, and (c) SPI-12 are also plotted. As a result of the regression analysis, SPI-1, SPI-3, SPI-9, SPI-4, and SPI-8 with standard coefficients of (0.711, − 0.215, 0.093, 0.092, and − 0.072) have been found to be the main input parameters that influence SPI drought index estimation for SPI-3 (Table 2). For SPI-6, the priority of the influential input parameters was SPI-1, SPI-6, SPI-7, SPI-3, and SPI-9 by providing absolute standard coefficients (0.742, − 0.242, 0.161, 0.110, and − 0.093), respectively. The influential input parameters were prioritized as SPI-1, SPI-

Table 3 MAE, RMSE, RAE, RRSE, and R for Machine learning algorithms-based models during the training and testing span

Machine learning algorithm	Training					Testing				
	MAE	RMSE	RAE	RRSE	R ²	MAE	RMSE	RAE	RRSE	R ²
(A) SPI-3										
RSS	0.540	0.685	71.07	71.19	0.756	0.522	0.678	85.99	86.79	0.462
RSS-M5P	0.523	0.679	68.86	70.53	0.714	0.497	0.682	81.88	87.22	0.507
RSS-RF	0.367	0.484	48.34	50.25	0.879	0.618	0.761	101.71	97.29	0.617
RSS-RT	0.505	0.615	66.51	63.89	0.805	0.563	0.726	92.61	92.87	0.353
(B) SPI-6										
RSS	0.427	0.567	52.35	55.14	0.870	0.570	0.759	88.04	90.27	0.365
RSS-M5P	0.405	0.559	49.71	54.33	0.840	0.452	0.717	69.76	85.24	0.402
RSS-RF	0.305	0.425	37.36	41.25	0.919	0.571	0.795	88.24	94.54	0.302
RSS-RT	0.353	0.494	43.33	47.98	0.883	0.557	0.781	86.01	92.90	0.290
(C) SPI-12										
RSS	0.251	0.386	30.72	36.14	0.937	0.347	0.428	65.87	67.68	0.796
RSS-M5P	0.236	0.371	28.88	34.74	0.938	0.294	0.377	55.79	59.57	0.783
RSS-RF	0.209	0.332	25.59	31.08	0.951	0.376	0.495	71.39	78.35	0.707
RSS-RT	0.261	0.428	31.93	40.13	0.917	0.399	0.473	75.67	74.76	0.757

2, and SPI-10, respectively, using absolute standard coefficients of (1.062, − 0.165, and − 0.104) for SPI-12.

2.5 Performance metrics and evaluation

During the period of this study, actual data was compared with modeled values. Statistical indicators have been used in evaluating the accuracy of developed hybrid Random SubSpace models, e.g., Root mean square error (RMSE), coefficient of determination (R²), relative absolute error (RAE), root relative squared error (RRSE) and mean absolute error (MAE) (Kushwaha et al. 2021; Elbeltagi et al. 2022b). All statistical indicators are defined as:

1. Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (SPI_A^i - SPI_P^i)^2} \tag{6}$$

2. Coefficient of determination (R²)

$$R^2 = \left[\frac{\sum_{i=1}^N (SPI_A^i - \overline{SPI_A})(SPI_P^i - \overline{SPI_P})}{\sqrt{\sum_{i=1}^N (SPI_A^i - \overline{SPI_A})^2} \sqrt{\sum_{i=1}^N (SPI_P^i - \overline{SPI_P})^2}} \right]^2 \tag{7}$$

3. Mean absolute error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |SPI_P^i - SPI_A^i| \tag{8}$$

4. Relative absolute error (RAE)

The RAE normalizes the total absolute error by

dividing it by the simple predictor’s total absolute error.

$$RAE = \left| \frac{SPI_A^i - SPI_P^i}{SPI_P^i} \right| \times 100 \tag{9}$$

5. Root relative squared error (RRSE)

The RRSE normalizes the overall squared error by dividing it by the total SE of the simple predictor. The error is reduced to the same dimensions as the quantity being predicted by calculating the square root of the RSE.

$$RRSE = \frac{\sqrt{\sum_{i=1}^N (SPI_P^i - SPI_A^i)^2}}{\sqrt{\sum_{i=1}^N (SPI_A^i - SPI^-)^2}} \tag{10}$$

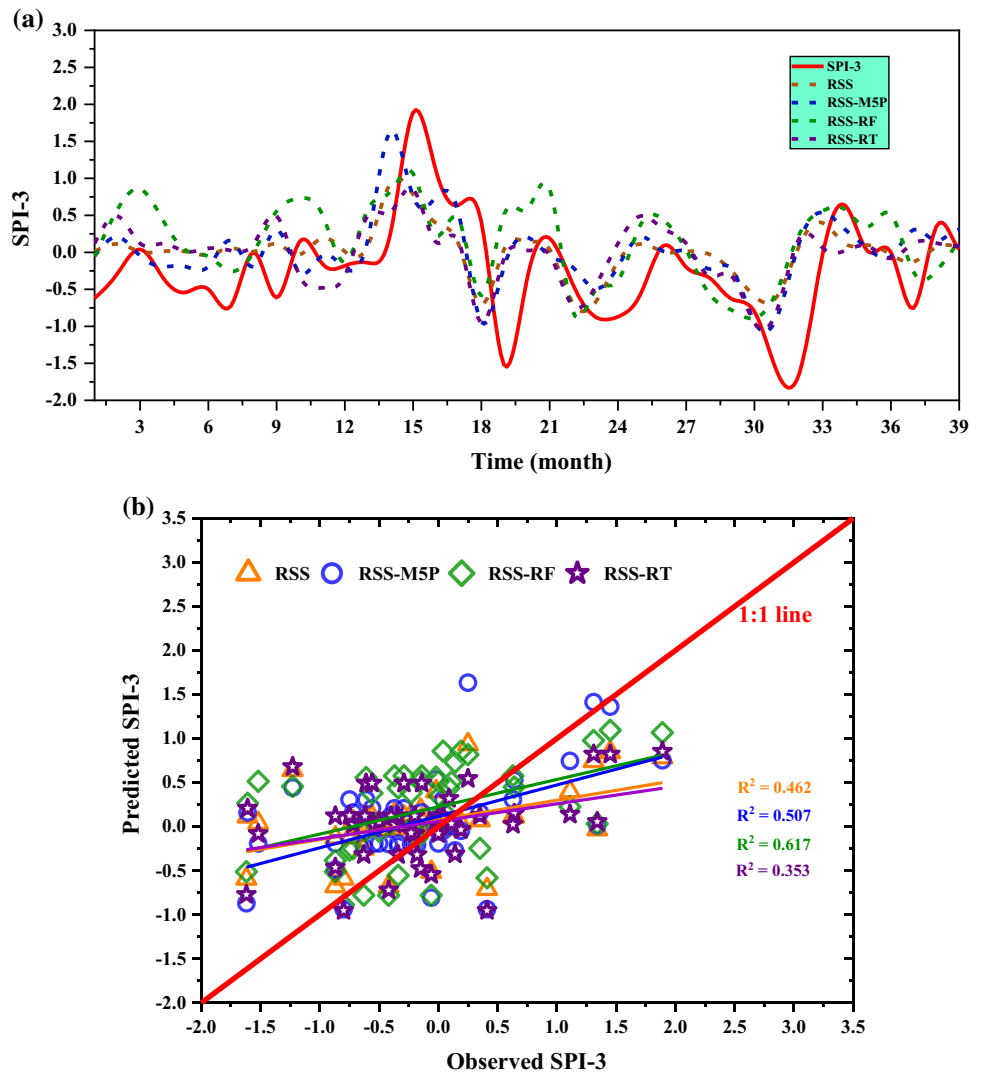
In which, SPI_A^i is an observed or actual value, SPI_P^i is simulated or forecasted value, $\overline{SPI_A}$ and $\overline{SPI_P}$ are the mean values of observed and forecasted samples, and N is the total number of data points.

3 Results

3.1 Evaluation machine learning models based on the best-selected subset models

Four machine learning methods (i.e., RSS, RSS-M5P, RSS-RF, and RSS-RT) were used to forecast the SPI at 3, 6, and 12 months at Jaisalmer district, Rajasthan, India. The

Fig. 5 Predicted and calculated SPI-3 values by RSS, RSS-M5P, RSS-RF and RSS-RT algorithms during testing phases **a** time series, and **b** scenario-scatter plot

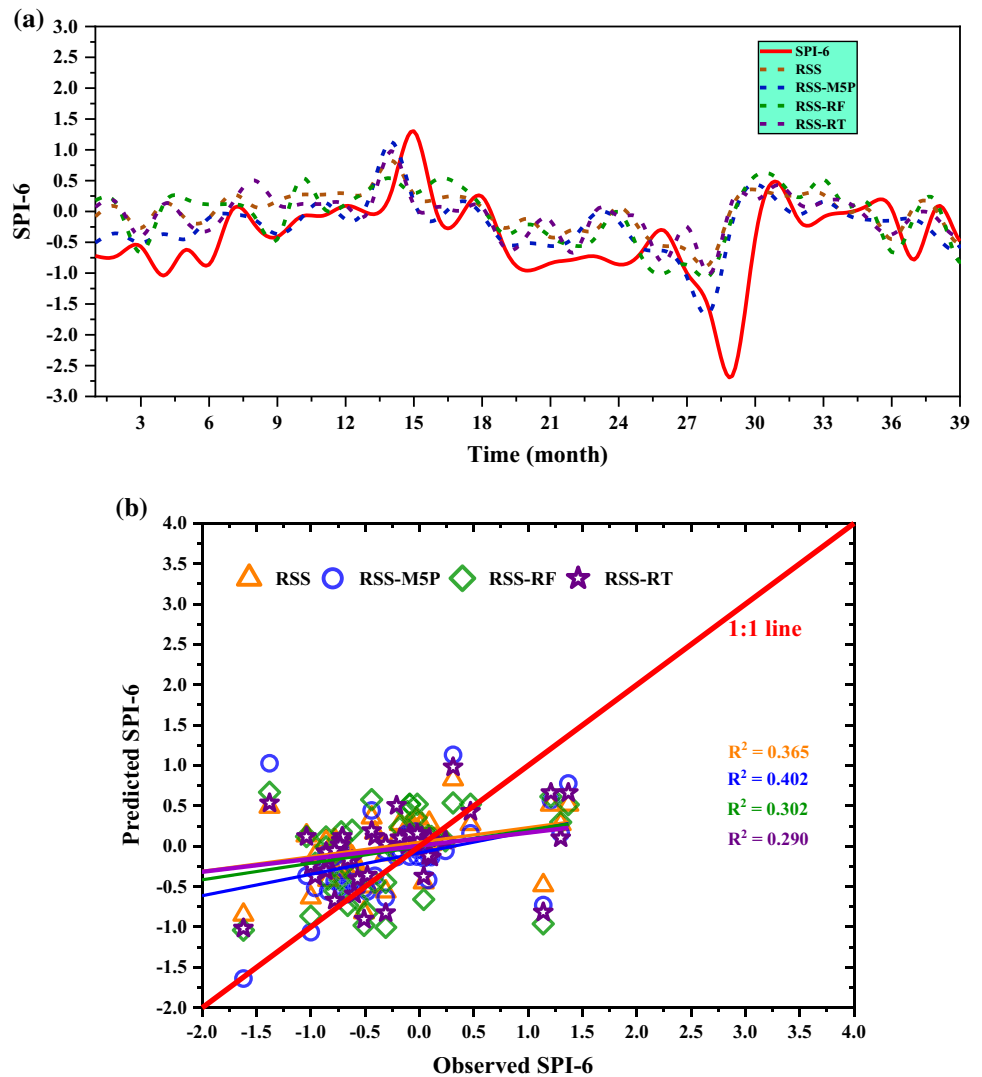


employed algorithms’ performances were evaluated and compared (i.e., MAE, RMSE, RAE, RRSE, and R^2). The model with the lowest MAE, RMSE, RAE, RRSE near zero, and R^2 near one is deemed to have the most accuracy in estimating the SPI. Table 3 shows the performance indices for machine learning algorithms-based models during the training and testing span. The best machine learning algorithm for each time-scales of SPI (i.e., SPI-3, SPI-6, and SPI-12) is displayed in blue row.

Results indicated that the RSS-RF model outperformed other algorithms during the training period for forecasting the SPI at 3, 6, and 12 months. It provided MAE = 0.367, RMSE = 0.484, RAE = 48.34, RRSE = 50.25, and $R^2 = 0.879$ for SPI-3. It gave MAE = 0.305, RMSE = 0.425, RAE = 37.36, RRSE = 41.25, and $R^2 = 0.919$ for SPI-6. And it has MAE = 0.209, RMSE = 0.332, RAE = 25.59, RRSE = 31.08, and $R^2 = 0.951$ for SPI-12. It was followed by RSS-RF, RSS-RT, and RSS algorithm, respectively. Among three months of the SPI predictive

models, SPI-12 has the highest performance, followed by SPI-6, and SPI-3, respectively, for the training period. The RSS-M5P algorithm outperformed the other implemented algorithms during testing. Therefore, it should consider RSS-M5P the best SPI prediction model. It provided MAE = 0.497, RMSE = 0.682, RAE = 81.88, RRSE = 87.22, and $R^2 = 0.507$ for SPI-3. It gave MAE = 0.452, RMSE = 0.717, RAE = 69.76, RRSE = 85.24, and $R^2 = 0.402$ for SPI-6. And it has MAE = 0.294, RMSE = 0.377, RAE = 55.79, RRSE = 59.57, and $R^2 = 0.783$ for SPI-12. Again, among three months of the SPI predictive models, SPI-12 has the highest performance, followed by SPI-6, and SPI-3, respectively, for the testing period. Figures 5, 6 and 7 present the predicted and calculated SPI-3, SPI-6, and SPI-12 values by four machine learning algorithms during testing phases: (a) time series and (b) scenario-scatter plot. Further comparative examination of models was done using the Taylor diagram (Fig. 8). According to standard deviation, correlation, and RMSE,

Fig. 6 Predicted and calculated SPI-6 values by RSS, RSS-M5P, RSS-RF and RSS-RT algorithms during testing phases **a** time series, and **b** scenario-scatter plot



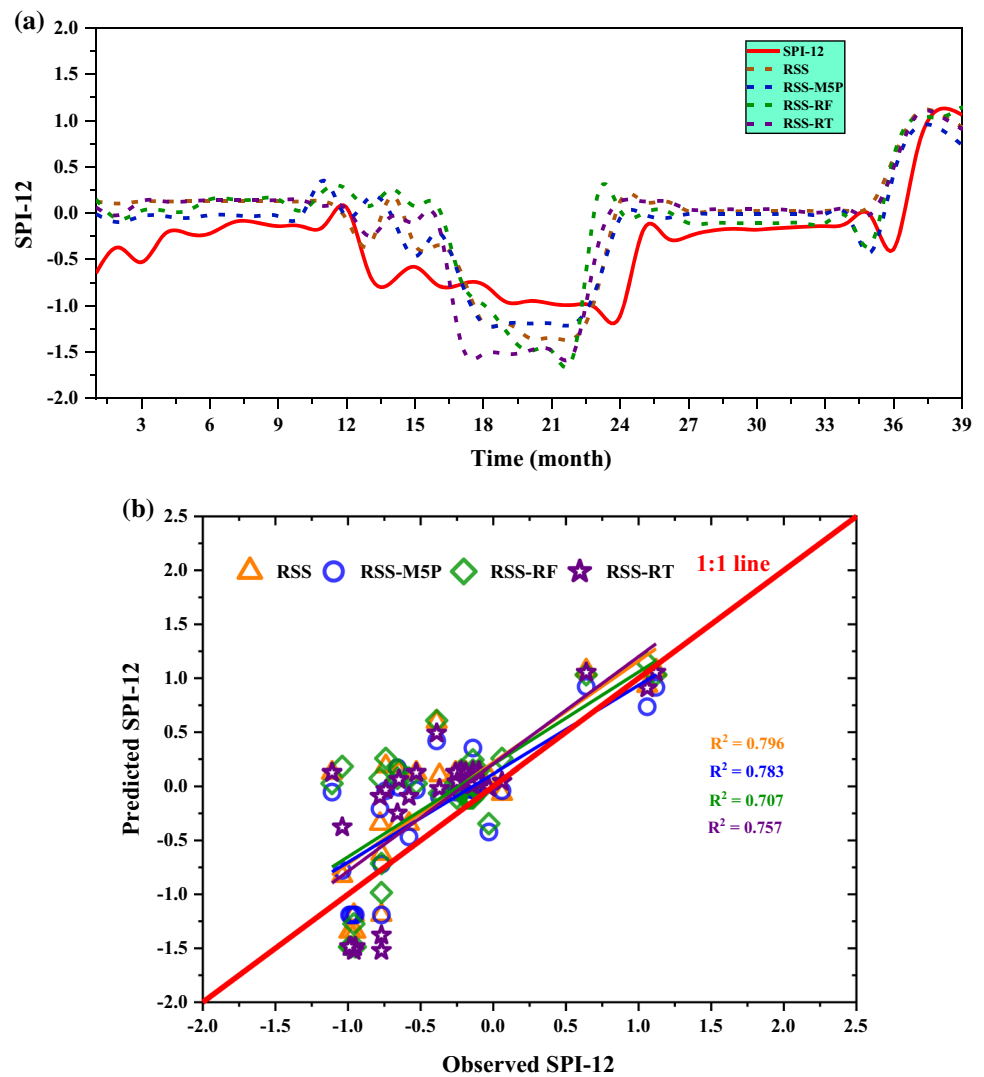
the RSS-M5P model matched the observed location the closest, while RSS, RSS-RF, and RSS-RT models further matched the location. In this analysis, RSS-M5P was the most effective model among the chosen models due to giving more generalized performance than RSS, RSS-RF, and RSS-RT algorithm. All the developed models could be observed as underestimating SPI-3 and SPI-6 while providing overestimation for SPI-12.

4 Discussion

The performance of hybrid meta-heuristics algorithms i.e. RSS, RSS-M5P, RSS-RF and RSS-RT was assessed for the multiscale prediction of SPI (i.e. 3, 6 and 12 months). The obtained results highlighted the potential of hybrid meta-heuristics algorithms in the prediction of monthly SPI especially for SPI-12 months. Figures 5, 6 and 7 represented the temporal variation between predicted and

calculated SPI values and their scatter plots for SPI-3, SPI-6, and SPI-12 months. In scatter plots, the regression line provided the high value of coefficient of determination (R^2) in respect of the RSS-M5P additive regression model under the both scenario for SPI-3, SPI-6, and SPI-12 months. As seen from the figures, the hybrid meta-heuristics algorithms performed better in longer time scale. Further comparison between algorithms using MAE and RMSE (Table 3) showed that M5P algorithms have improved the performance of the RSS structure as it has lower value of MAE and RMSE. The Taylor Diagram (Fig. 8) showed the more comparable depiction of models performance in prediction of monthly SPI values. The developed RSS-RF model was located furthest and RSS-M5P model was located nearest to the observed point based on the standard deviation, correlation, and RMSE for all monthly SPI prediction. This showed RSS-M5P algorithm has higher accuracy in prediction of monthly SPI as compared to RSS alone and other developed hybrid algorithms.

Fig. 7 Predicted and calculated SPI-12 values by RSS, RSS-M5P, RSS-RF and RSS-RT algorithms during testing phases **a** time series, and **b** scenario-scatter plot



Our findings were also compared with other recent studies conducted across different regions, such as Bangladesh, Ethiopia, India, and Iran. Considering both training and testing periods, the long-term SPI predictive model makes more accurate predictions than the short-term SPI predictive models. This present finding agree with the study by (Aghelpour and Varshavian 2021; Malik et al. 2021; Yaseen et al. 2021). This reflects that long-term precipitation patterns vary less than short and medium-term precipitation patterns (Belayneh and Adamowski 2013). Furthermore, monsoon months are more vulnerable than other seasons, with June showing greater susceptibility to severe drought. September displays greater susceptibility to extreme droughts. Using the RSS-RF model for each time-scale of SPI (i.e., SPI-3SPI-6 and SPI-12) gave the best performance among the selected models during the training period, while it was not the best for the testing period. It is in line with the study by Ditthakit et al. (2021), who applied the RF method for estimating GR2M model

parameters in an ungauged basin. Yaseen et al. (2021) investigated the capability of machine learning (ML) random forest (RF), minimum probability machine regression (MPMR), M5 Tree (M5tree), extreme learning machine (ELM), and online sequential-ELM (OSELM) in predicting (SPI) at four-month horizons (i.e., 1, 3, 6 and 12) in Bangladesh. Study found that ELM was the best model for predicting 3, 6, and 12-month SPI, while RF showed the best performance for 1- month SPI prediction. Belayneh et al. (2016) applied three machine learning techniques, i.e., artificial neural networks (ANNs), support vector regression (SVR), and coupled wavelet-ANNs (WA-ANN). They concluded that WA-ANN gave the best model performance for forecasting SPI 3 (3-month SPI) and SPI 6 (6-month SPI) in the Awash River Basin in Ethiopia. Aghelpour and Varshavian (2021) applied a hybrid model of MLP Neural Network and the Imperialistic Competitive Algorithm (MLP-ICA) for forecasting Multivariate Standardized Precipitation Index (MSPI) in Iran. They pointed

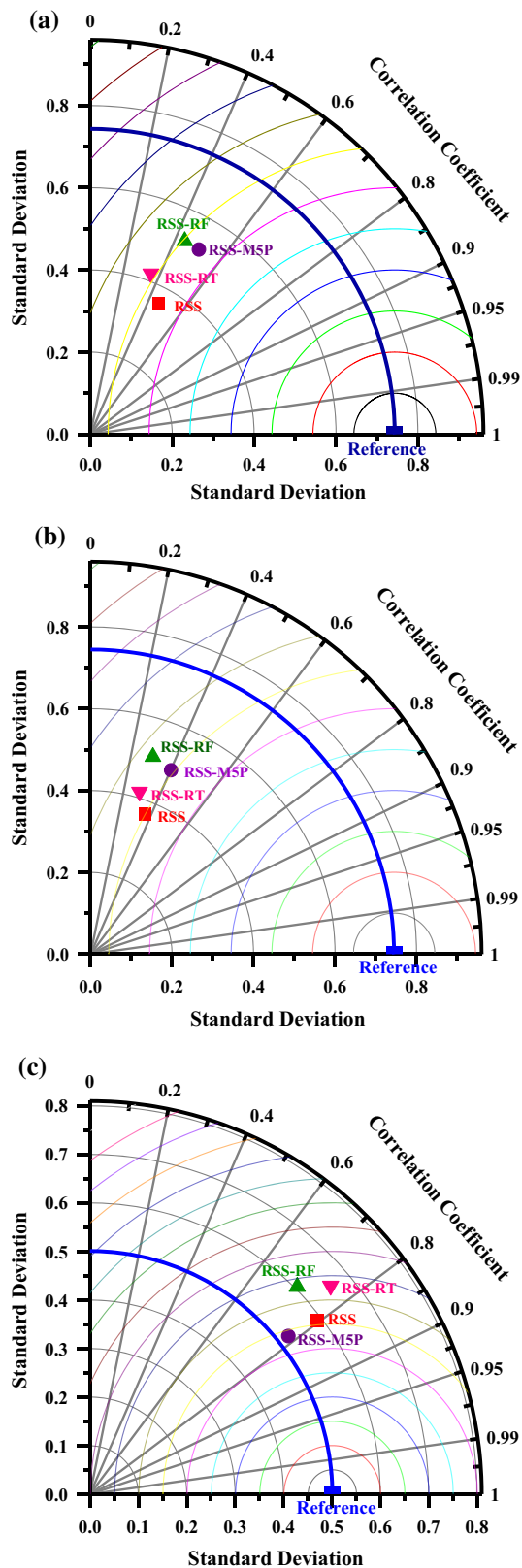


Fig. 8 Taylor diagrams of RSS, RSS-M5P, RSS-RF and RSS-RT during testing span at selected station for **a** SPI-3, **b** SPI-6, and **c** SPI-12

out the proposed models could forecast MSPI for a longer time horizon (i.e., 12–24 and 24–48 month MSPI) were better than the shorter time horizon (i.e., 3–6, 6–12, and 3–12 month MSPI). The present study highlighted the potential of hybrid meta-heuristics algorithms in the prediction of multi scale SPI droughts. However, uncertainties related to datasets, methods, scenarios, models with a large number of algorithms parameters, search space and optimization process becomes more difficult. Therefore, further research should focus on minimizing these uncertainties and improving optimization performance. Moreover, since only one station was selected in this study, the applicability of the proposed RSS-based hybrid models may be validated at different locations varying in agro-climatic conditions under different scenarios to draw a generalized conclusion.

5 Conclusion

Assessment of drought is one of the most important tasks in the present condition as it leads to several adverse effects on the soil–water–atmosphere cycles of the earth systems across the different climates of the world. There's exists several techniques and machine learning methods in the literature for quantifying drought, However, based on SPI, this is one of the unique study that compares and contrast the relative role of meta-heuristic models such as RSS alone and its hybridization with other algorithms. Drought is among the most global costly threats to ecosystems, especially in regions with diverse climatic patterns. Drought occurs, as a spatio-temporal phenomenon, as a result of a decrease in the amounts of precipitation below average for a specific period sufficient to cause environmental risks. In the present study, four machine learning models (i.e., RSS, RSS-M5P, RSS-RF, and RSS-RT) are applied to the SPI (Standard Precipitation Index) evaluation in Rajasthan climate conditions to create a probabilistic framework for drought situations. The maximum drought periods and the corresponding durations have been identified in the study location, and the results confirm that there have been some severe drought events in the past. The different SPI timescales (SPI-3, SPI-6, and SPI-12) present distinct drought periods and intensities that are vital to a seasonal drought analysis. In the training stages of all the climates, AI models developed using the RSS-M5P model had the highest efficiency and was followed by RSS, RSS-RF, and RSS-RT models. According to SPI-3 results, monsoon months are more vulnerable than other seasons, with June showing greater susceptibility to severe drought. September displays greater susceptibility to extreme droughts. The SPI-12 model exhibited the highest performance among models developed for three months,

followed by SPI-6 and SPI-3. In general, all the models provided an underestimation of SPI-3 and SPI-6, while overestimating SPI-12. Results revealed that RSS-M5P model performed well in capturing the monthly trend of SPI and it has high values coefficient of determination (0.507–0.783) and lower values of MAE (0.294–0.497) and RMSE (0.377–0.682) for prediction of multi scale SPI (i.e., 3, 6, and 9 months), under the testing period. The occurrence of drought is affected by low precipitation, high fluctuations in the average rainfall, and climate change, particularly as a result of regional and global warming. As such, it is essential to develop drought management policies and effectively implement these policies with the assistance and support of the government and private organizations. We could use the results of our study to understand the water availability for the entire year in advance with the help of three months of SPI data, and that in turn could be used to create an efficient climate-smart agriculture strategy that has global economic implications.

Acknowledgements The authors are also thankful to the anonymous reviewers for their valuable comments and suggestions to improve this manuscript further.

Authors contributions AE, MK and NLK: Conceptualization, Methodology, Formal analysis, Software, Writing-Original draft preparation. NLK, DKV, MK, CBP, PD, and AS: Visualization, Comments and Revisions recommendations, Writing-Reviewing and Editing. AE, CBP, PD: Supervision, Comments and Revisions Recommendations, Writing-Reviewing and Editing.

Funding No funding was received for conducting this study.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval All authors comply with the guidelines of the journal *Stochastic Environmental Research and Risk Assessment*.

Consent to participate All authors agreed to participate in this study.

Consent to publication All authors agreed to the publication of this manuscript.

References



- Abd-Elaty I, Kushwaha NL, Grismer ME et al (2022a) Cost-effective management measures for coastal aquifers affected by saltwater intrusion and climate change. *Sci Total Environ* 836:155656. <https://doi.org/10.1016/j.scitotenv.2022.155656>
- Abd-Elaty I, Shoshah H, Zeleňáková M et al (2022b) Forecasting of flash floods peak flow for environmental hazards and water harvesting in desert area of El-Qaa Plain, Sinai. *Int J Environ Res Public Health* 19:6049. <https://doi.org/10.3390/ijerph19106049>
- Aghelpour P, Varshavian V (2021) Forecasting different types of droughts simultaneously using multivariate standardized precipitation index (MSPI), MLP neural network, and imperialistic competitive algorithm (ICA). *Complexity* 2021:e6610228. <https://doi.org/10.1155/2021/6610228>
- Ali M, Deo RC, Maraseni T, Downs NJ (2019) Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *J Hydrol* 576:164–184. <https://doi.org/10.1016/j.jhydrol.2019.06.032>
- Alizamir M, Kisi O, Muhammad Adnan R, Kuriqi A (2020) Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. *Acta Geophys* 68:1113–1126. <https://doi.org/10.1007/s11600-020-00446-9>
- Anand J, Gosain AK, Khosa R, Srinivasan R (2018) Regional scale hydrologic modeling for prediction of water balance, analysis of trends in streamflow and variations in streamflow: the case study of the Ganga River basin. *J Hydrol Reg Stud* 16:32–53. <https://doi.org/10.1016/j.ejrh.2018.02.007>
- Behnood A, Behnood V, Modiri Gharehveran M, Alyamac KE (2017) Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Constr Build Mater* 142:199–207. <https://doi.org/10.1016/j.conbuildmat.2017.03.061>
- Belayneh A, Adamowski J (2013) Drought forecasting using new machine learning methods. *J Water Land Dev* 18:3–12. <https://doi.org/10.2478/jwld-2013>
- 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>
- Biau G, Scornet E (2015) A random forest guided tour. *arXiv:151105741 [math, stat]*
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen F, Mitchell K, Schaake J et al (1996) Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J Geophys Res Atmos* 101:7251–7268. <https://doi.org/10.1029/95JD02165>
- 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>
- Dong X, Yu Z, Cao W et al (2020) A survey on ensemble learning. *Front Comput Sci* 14:241–258. <https://doi.org/10.1007/s11704-019-8208-z>
- Elbeltagi A, Aslam MR, Malik A et al (2020) The impact of climate changes on the water footprint of wheat and maize production in the Nile Delta, Egypt. *Sci Total Environ* 743:140770. <https://doi.org/10.1016/j.scitotenv.2020.140770>
- Elbeltagi A, Aslam MR, Mokhtar A et al (2021) Spatial and temporal variability analysis of green and blue evapotranspiration of wheat in the Egyptian Nile Delta from 1997 to 2017. *J Hydrol* 594:125662. <https://doi.org/10.1016/j.jhydrol.2020.125662>
- Elbeltagi A, Di Nunno F, Kushwaha NL et al (2022a) River flow rate prediction in the Des Moines watershed (Iowa, USA): a machine learning approach. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-022-02228-9>
- Elbeltagi A, Kushwaha NL, Rajput J et al (2022b) Modelling daily reference evapotranspiration based on stacking hybridization of ANN with meta-heuristic algorithms under diverse agro-climatic conditions. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-022-02196-0>

- Healey SP, Cohen WB, Yang Z et al (2018) Mapping forest change using stacked generalization: an ensemble approach. *Remote Sens Environ* 204:717–728. <https://doi.org/10.1016/j.rse.2017.09.029>
- Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20:832–844. <https://doi.org/10.1109/34.709601>
- Kalmegh S (2015) Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news. *Int J Innov Sci Eng Technol* 2:438–446
- Khosravi K, Daggupati P, Alami MT et al (2019) Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: a case study in Iraq. *Comput Electron Agric*. <https://doi.org/10.1016/j.compag.2019.105041>
- Kuriqi A, Ali R, Pham QB et al (2020a) Seasonality shift and streamflow flow variability trends in central India. *Acta Geophys* 68:1461–1475. <https://doi.org/10.1007/s11600-020-00475-4>
- Kuriqi A, Pinheiro AN, Sordo-Ward A, Garrote L (2020b) Water-energy-ecosystem nexus: balancing competing interests at a run-of-river hydropower plant coupling a hydrologic–ecohydraulic approach. *Energy Convers Manage* 223:113267. <https://doi.org/10.1016/j.enconman.2020.113267>
- Kushwaha NL, Bhardwaj A, Verma VK (2016) Hydrologic response of Takarla-Ballowal watershed in Shivalik foot-hills based on morphometric analysis using remote sensing and GIS. *J Indian Water Resour Soc* 36:17–25
- Kushwaha NL, Rajput J, Elbeltagi A et al (2021) Data intelligence model and meta-heuristic algorithms-based pan evaporation modelling in two different agro-climatic zones: a case study from Northern India. *Atmosphere* 12:1654. <https://doi.org/10.3390/atmos12121654>
- Kushwaha N, Elbeltagi A, Mehan S et al (2022a) Comparative study on morphometric analysis and RUSLE-based approaches for micro-watershed prioritization using remote sensing and GIS. *Arab J Geosci* 15:564. <https://doi.org/10.1007/s12517-022-09837-2>
- Kushwaha NL, Rajput J, Shirsath PB et al (2022) Seasonal climate forecasts (SCFs) based risk management strategies: a case study of rainfed rice cultivation in India. *J Agrometeorol* 24:10–17. <https://doi.org/10.54386/jam.v24i1.775>
- Malik A, Kumar A, Rai P, Kuriqi A (2021) Prediction of multi-scalar standardized precipitation index by using artificial intelligence and regression models. *Climate* 9:28. <https://doi.org/10.3390/cli9020028>
- Mao J, Fu W, Shi X et al (2015) Disentangling climatic and anthropogenic controls on global terrestrial evapotranspiration trends. *Environ Res Lett* 10:094008. <https://doi.org/10.1088/1748-9326/10/9/094008>
- McKee T, Doesken N, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th conference on applied climatology*. Anaheim, CA, USA, pp 179–184
- Melesse AM, Khosravi K, Tiefenbacher JP et al (2020) River water salinity prediction using hybrid machine learning models. *Water* 12:2951
- Misra S, Li H (2020) Chapter 9—noninvasive fracture characterization based on the classification of sonic wave travel times. In: Misra S, Li H, He J (eds) *Machine learning for subsurface characterization*. Gulf Professional Publishing, Houston, pp 243–287
- Mosavi A, Shirzadi A, Choubin B et al (2020) Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.3014816>
- Muhammad Adnan R, Chen Z, Yuan X et al (2020) Reference evapotranspiration modeling using new heuristic methods. *Entropy* 22:547. <https://doi.org/10.3390/e22050547>
- Nhu V-H, Shahabi H, Nohani E et al (2020) Daily water level prediction of Zrebar Lake (Iran): a comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. *ISPRS Int J Geo Inf* 9:479. <https://doi.org/10.3390/ijgi9080479>
- Pham BT, Prakash I, Tien Bui D (2018) Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* 303:256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>
- Poomima S, Pushpalatha M (2019) Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Comput* 23:8399–8412. <https://doi.org/10.1007/s00500-019-04120-1>
- Quinlan JR (1992) *Learning with continuous classes*. World Scientific, Singapore, pp 343–348
- Rahman M, Chen N, Elbeltagi A et al (2021) Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. *J Environ Manag* 295:113086. <https://doi.org/10.1016/j.jenvman.2021.113086>
- Rodell M, Velicogna I, Famiglietti JS (2009) Satellite-based estimates of groundwater depletion in India. *Nature* 460:999–1002. <https://doi.org/10.1038/nature08238>
- Sammut C, Webb GI (2017) Random subspace method. In: Sammut C, Webb GI (eds) *Encyclopedia of machine learning and data mining*. Springer US, Boston, pp 1055–1055
- Sikora R, Al-Laymoun O, Sikora R, Al-Laymoun O (2015) A modified stacking ensemble machine learning algorithm using genetic algorithms through big data analytics. In: Tavana M (ed) *Handbook of research on organizational transformations through big data analytics*. IGI Global, Hershey, pp 43–53
- Suwal N, Kuriqi A, Huang X et al (2020) Environmental flows assessment in nepal: the case of Kaligandaki River. *Sustainability* 12:8766. <https://doi.org/10.3390/su12218766>
- Tao H, Diop L, Bodian A et al (2018) Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso. *Agric Water Manag* 208:140–151. <https://doi.org/10.1016/j.agwat.2018.06.018>
- Thomas T, Jaiswal RK, Nayak PC, Ghosh NC (2015) Comprehensive evaluation of the changing drought characteristics in Bundelkhand region of Central India. *Meteorol Atmos Phys* 127:163–182. <https://doi.org/10.1007/s00703-014-0361-1>
- Tigkas D, Vangelis H, Tsakiris G (2015) DrinC: a software for drought analysis based on drought indices. *Earth Sci Inform* 8:697–709. <https://doi.org/10.1007/s12145-014-0178-y>
- Vicente-Serrano SM, Quiring SM, Peña-Gallardo M et al (2020) A review of environmental droughts: increased risk under global warming? *Earth Sci Rev* 201:102953. <https://doi.org/10.1016/j.earscirev.2019.102953>
- Wang K, Dickinson RE (2012) A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. *Rev Geophys*. <https://doi.org/10.1029/2011RG000373>
- Wang K, Dickinson RE, Wild M, Liang S (2010) Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. *J Geophys Res Atmos*. <https://doi.org/10.1029/2009JD013671>
- Wang Y, Witten IH (1996) Induction of model trees for predicting continuous classes
- Wolpert DH (1992) Stacked generalization. *Neural Netw* 5:241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)

- Yang Y, Long D, Shang S (2013) Remote estimation of terrestrial evapotranspiration without using meteorological data. *Geophys Res Lett* 40:3026–3030. <https://doi.org/10.1002/grl.50450>
- Yaseen ZM, Ali M, Sharafati A et al (2021) Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh. *Sci Rep* 11:3435. <https://doi.org/10.1038/s41598-021-82977-9>
- Zeng Z, Peng L, Piao S (2018) Response of terrestrial evapotranspiration to Earth's greening. *Curr Opin Environ Sustain* 33:9–25. <https://doi.org/10.1016/j.cosust.2018.03.001>
- Zhou Z-H (2009) Ensemble learning. In: Li SZ, Jain A (eds) *Encyclopedia of biometrics*. Springer US, Boston, pp 270–273

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

Authors and Affiliations

Ahmed Elbeltagi¹ · Manish Kumar² · N. L. Kushwaha³  · Chaitanya B. Pande⁴ · Pakorn Ditthakit⁵ · Dinesh Kumar Vishwakarma⁶  · A. Subeesh⁷

✉ N. L. Kushwaha
nand.kushwaha@icar.gov.in

Ahmed Elbeltagi
ahmedelbeltagy81@mans.edu.eg

Manish Kumar
manishgbpuat@yahoo.com

Chaitanya B. Pande
chaitanay45@gmail.com

Pakorn Ditthakit
dpakorn@mail.wu.ac.th

Dinesh Kumar Vishwakarma
dinesh.vishwakarma4820@gmail.com

A. Subeesh
subeesh.a@icar.gov.in

² College of Agricultural Engineering and Technology, Dr. R.P.C.A.U, Pusa, Bihar 848125, India

³ Division of Agricultural Engineering, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

⁴ National Institute of Rural Development & Panchayati Raj, Hyderabad, India

⁵ Center of Excellence in Sustainable Disaster Management (CESDM), School of Engineering and Technology, Walailak University, 222, Thaiburi, Thasala, Nakhon Si Thammarat 80160, Thailand

⁶ Department of Irrigation and Drainage Engineering, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand 263145, India

⁷ ICAR-Central Institute of Agricultural Engineering (CIAE), Bhopal, Madhya Pradesh, India

¹ Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt