Designing a decomposition-based multi-phase pre-processing strategy coupled with EDBi-LSTM deep learning approach for sediment load forecasting

Mehdi Jamei [a,b,h,*], Mumtaz Ali [c,g,h], Anurag Malik [d], Priya Rai [e], Masoud Karbasi [f], Aitazaz A. Farooque [g,h], Zaher Mundher Yaseen [i,j]

[a] Faculty of Engineering, Shohadaye Hoveizeh Campus of Technology, Shahid Chamran University of Ahvaz, Dashte Azadegan, Iran

[b] New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq

[c] UniSQ College, University of Southern Queensland, QLD 4305, Australia

[d] Punjab Agricultural University, Regional Research Station, Bathinda-151001, Punjab, India

[e] Department of Soil and Water Conservation Engineering, College of Technology, G.B. Pant University of Agriculture and Technology, Pantnagar-263145, Uttarakhand, India

[f] Water Engineering Department, Faculty of Agriculture, University of Zanjan, Zanjan, Iran

[g] Faculty of Sustainable Design Engineering, University of Prince Edward Island, Charlottetown, PE, Canada

[h] Canadian Centre for Climate Change and Adaptation, University of Prince Edward Island, St Peters, PE, Canada

[i] Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

[j] Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

ABSTRACT

Forecasting accurately suspended sediment load (SSL) in the basin is one of the most critical issues for river engineering, environment, and water resources management which effectively reduces flood damages. In this study, a new multi-criteria hybrid expert system comprised of empirical wavelet decomposition (EWT) integrated with Encoder-Decoder Bidirectional long short-term memory (EDBi-LSTM), supported by five feature selection (FS) methods was developed for the first time to forecast daily SSL at two study sites (Bamini and Ashti) of Godavari river basin, India. The employed FS schemes are including Boruta-Random forest (BRF), simulated annealing (SA), Relief algorithm, Ridge regression (RR), and Mutual information (MI) where the BRF coupled with EWT and EDBi-LSTM (i.e., E_{WT}-E_{DBi}-LSTM-Boruta) is identified as the main forecasting paradigm. Here the original SSL signals in the monsoon season (2001–2015) as the only input information were considered to forecast SSL events at a daily time scale in both study zones. The SSL signals were decomposed using the EWT technique considering the significant antecedent time-lagged inputs based on partial auto-correlation function (PACF). In the next stage, five FS strategies were addressed to specify the significant sub-sequences to reduce computational cost and enhance forecasting accuracy. Besides, the extreme gradient boosting (XGB) approach was implemented to compare the potential of the hybrid EDBi-LSTM and standalone counterpart models for both study sites. According to several goodness-of-fit indices and validation tools, the outcomes at the Bamini and Ashti sites demonstrated that the E_{WT}-E_{DBi}-LSTM-Boruta as the main model, achieved the best accuracy, followed by E_{WT}-X_{GB}-Boruta, E_{WT}-E_{DBi}-LSTM-SA, and E_{WT}-X_{GB}-SA, respectively. Comparing all the hybrid models showed that the BRF, SA, and RR strategies performed better in integration with machine learning (ML) models.

1. Introduction

River systems are frequently regulated for numerous reasons, such as flood control, navigation, irrigation, water supply, hydropower generation, and others (Evaristo and McDonnell, 2019). River systems are also employed to dispose of wastewater and rainwater from textiles, pharmaceuticals, and mills (Awadh and Yousif, 2021; Bhagat et al., 2018; Bhagat and Tiyasha, 2013; Yaseen et al., 2019b). Water treatment using various civil structures and hydrological control systems is

Nomenclature		mm	Millimetre
		NSE	Nash–Sutcliffe efficiency
AI	Artificial intelligence	^{o}C	Degree Celsius
ANN	Artificial neural network	OOB	Out-of-bag
BRF	Boruta-Random forest	PACF	Partial auto-correlation function
DT	Decision tree	R	Correlation coefficient
EDBi-LSTM	Encoder-Decoder Bidirectional long short-term memory	RAE	Relative absolute error
EWT	Empirical wavelet transform	REB	Residual error band
FL	Fuzzy logic	RFS	Relief feature selection
FS	Feature selection	RLU	Rectified linear unit
g/L	Gram per litre	RMSE	Root mean square error
GBT	Gradient boosting tree	RNN	Recurrent neural network
GRB	Godavari river basin	RR	Ridge regression
KGE	Kling-Gupta efficiency	SA	Simulated annealing
km	Kilometre	SRC	Sediment rating curve
LSTM	Long short-term memory	SSL	Suspended sediment load
m	Metre	SST	Suspended sediment transport
MDA	Mean decrease accuracy	SVM	Support vector machines
MI	Mutual information	U$_{95\%}$	Uncertainty coefficient with 95% confidence level
ML	Machine learning	XGB	Extreme gradient boosting

typically required to effectively exploit a river system's resources and mitigate the associated risks (Kisi et al., 2019; Yaseen et al., 2019a). An excessive amount of silt hampers the operation of hydraulic machinery in the water (Betrie et al., 2011; Walling and Collins, 2008). High sediment concentrations also affect the general quality of water (Tao et al., 2019). As a result, resolving such issues necessitates predicting suspended loads which account for approximately 95% of total sediment loads (Simons and Şentürk, 1992). Considering the complex geometry of a river system which governs flow turbulence structure and water velocity and controls the sediment-carrying capacity of water that enters the river system (Armanini et al., 2015), the development of a robust model for the prediction of suspended sediment transport (SST) is still a problem.

Sediment transportation is on of the major contributor in enhancing environmental pollution, and tracking the erratic pattern of sediment transport is still lacking due to the integrated effects of known and unknown parameters having complex behavior that influence the transport process (Shojaeezadeh et al., 2018). There are no accurate and complete theoretical equations that can capture the two-phase concepts of fluid and SST, hydrologic research has traditionally depended on experimental-based similitude analysis. Sediment transportation in a river basin is related to fluid-sediment interaction and other complicated phenomena. In addition, the properties of flow and sediment are considered the essential physical parameters of sediment transport. Turbulence flow keeps sediment suspended due to threshold sediment speed. From the initial detachment of sediment particles through the arrival of sediment at a place of deposition, a combination of highly non-linear and interacting mechanisms contribute to this complexity at every stage of the transportation process. To calculate sediment transport variables, various experimental and analytical techniques have been developed (Li and Li, 2018). The geometric boundary and its water flow resistance, sediment transport rate, and sediment mass conservation have been described using techniques. The main aim of this paper is to apply the important artificial intelligence (AI)-based models developed to model suspended sediment in a river system. After studying the relevant literature on SST, the AI-based models are discussed in the next section.

The operation of dams, canals, and diversions requires accurate prediction of suspended sediment levels in water bodies (Azamathulla et al., 2012; Cigizoglu, 2004; Suif et al., 2016). Erosion and sediment transport are the climate-induced main cutting edge environmental and hydrological research challenges for decades.; therefore, many studies have been focused on river sediments and their impact on global surface water resource consumption (Greig et al., 2005; Malagó et al., 2017; Sinha et al., 2019). Deforestation and overgrazing contribute in eroding the soil surface and affect the sediment transport dynamics in river basin systems. Considering the drawback of deterministic physical and analytical models, the stochastic nature of river flow, initial and boundary conditions, and non-stationarity, have led to the development of AI-based models that can handle nonlinear relationships between water flow and environmental parameters for easy prediction of river sediment load. Given the numerous impacting factors, we must approach suspended sediment prediction as a complicated and nonlinear process for this article (Bhagat et al., 2020). These factors can be generally divided into four categories based on their origin (Tao et al., 2021a): (i) meteorological origin, (ii) hydrological origin, (iii) geological origin, and (iv) watershed geomorphology. This classification can inform our understanding of the causal inference between SST and its related predictors.

Sediment transport is still a trending topic due to its interdisciplinary nature with various approaches developed for its prediction in rivers (Williams and Berndt, 1976). Physical or deterministic methods can be time-consuming and inaccurate as they may incorporate initial or boundary conditions that differ widely for different watersheds (Afan et al., 2016). However, the nonlinear and intricate nature of SST in river bodies provides those such approaches can also be over-parameterized. Although such spatial data can be extracted from satellite or other sources (remotely sensed), it is not easy to obtain them at the correct level of initial condition for diverse watersheds, and as such, they may require calibration before being utilized as input for a physical model (Akay et al., 2008; Shamaei and Kaedi, 2016). Hence, understanding the physical processes involved in SST to water bodies is critical for the practical deployment and implementation of sediment mitigation techniques (Adams et al., 2018; Sadeghi and Singh, 2017). The problems with active storage and estimating the life of a reservoir have been documented in studies (De Vente et al., 2005). Reservoir sedimentation is still a global challenge that, if solved, might have global implications. Another issue is that reservoir sedimentation is a complicated transport mechanism in general (Verstraeten et al., 2003). The simulation of the behavior of SSL in rivers has traditionally relied on either basic statistical or numerical models (such as finite-difference approaches and sediment rating curves (SRC)) (Nguyen et al., 2009; Walling, 1977).
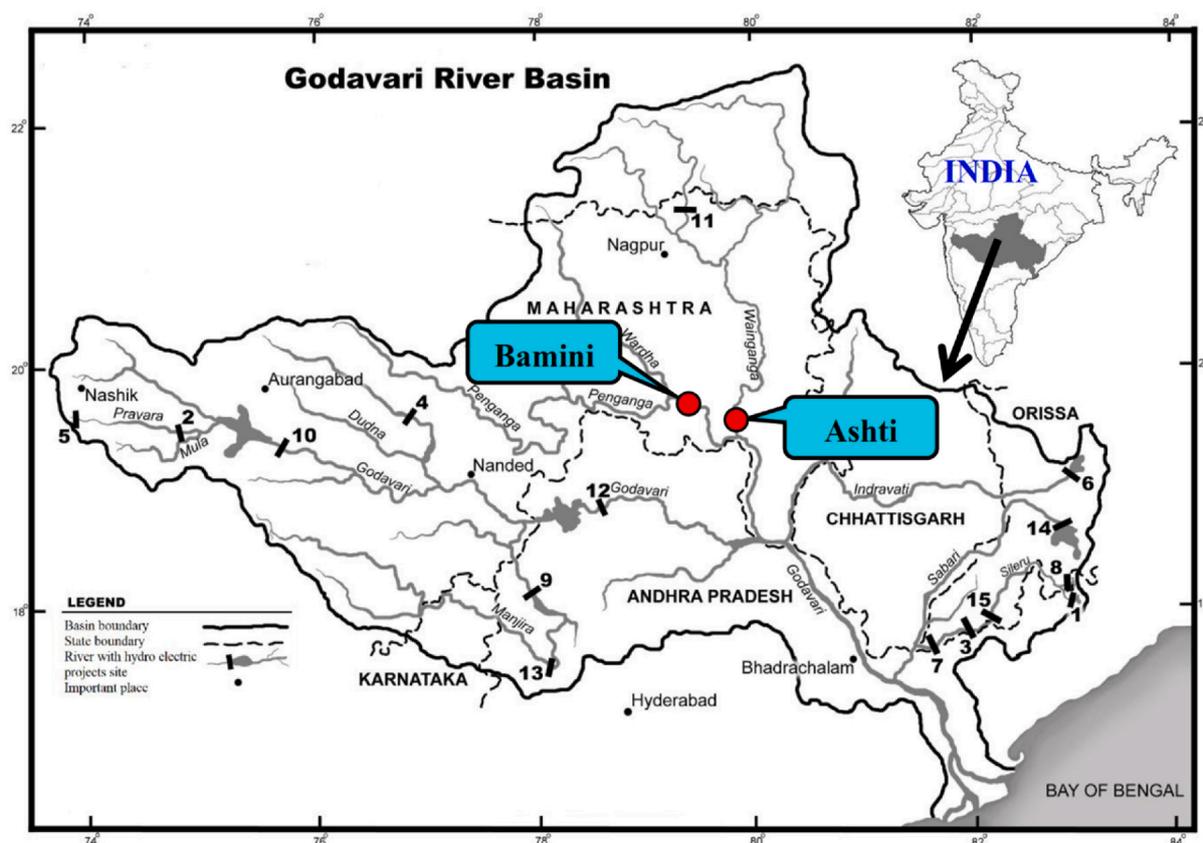
Fig. 1. The location map of study synaptic sites on GRB, India.

However, AI or machine learning (ML) models have recently emerged as approaches that operate by combining data mining approaches and soft computing methods to produce good performances, particularly in simulating nonlinear systems associated with hydrological processes in a bid to address water resources problems (Khosravi et al., 2019; Zounemat-Kermani et al., 2018). Given recent advances in AI models and their applications in predicting sediment transport in river basins, it is vital to establish a fresh ground to build on past and current progress on AI-based models. Researchers have consistently (Baniya et al., 2019; Fadaee et al., 2020; Sharafati et al., 2021; Shiri and Kişi, 2012) used AI-based models for SSL prediction. For example, the prediction of SSL use in artificial neural networks (ANN) in Cambodia has been reported (Melesse et al., 2011). The implemented model's capability was compared to that of some non-linear approaches, and the results showed satisfactory performance of the ANN model. To evaluate SSL in the Kaoping river basin, researchers used power regression, linear regression, ANN, and support vector machines (SVM) as the prediction models (Chiang and Tsai, 2011). The results demonstrated that the SVM outperformed the ANN and other regression techniques used in the study. At Kasol, India, ANN, fuzzy logic (FL), and decision tree (DT) approaches like M5 and RepTree were employed to predict suspended sediment (SS) (Senthil Kumar et al., 2012). According to the findings, the M5 strategy performed better than the other methods used. The performance of the M5 DT and wavelet regression model-based model was compared to the performance of ANN models, and the performance of the M5 DT and wavelet regression model-based strategy was found to be better than that of ANN (Goyal, 2014). Another study reported the use of ANN to estimate weekly sediment load. The determined and predicted sediment load values were correlated with the results, so the study recommended the ANN-based model for reliable SSL prediction (Hassan et al., 2015). A two-stage modeling technique was used to deal with the Spatiotemporal variance associated with SSL (Nourani et al., 2016). The

strategy analysis revealed that combining SVM and spatial statistics techniques gave better and more accurate SSL simulation and prediction. More AI models for predicting sediment transport in rivers have been developed and their potential has been validated (Buyukyildiz and Kumcu, 2017; Hassanpour et al., 2019; Khosravi et al., 2018).

Even though AI models for modeling SSL transport in rivers have been implemented in several ways, the introduced approaches have several drawbacks, such as local minima entrapment, internal parameters modification, model flexibility, etc. Furthermore, the stochasticity of the examined river's "catchment" can add to the difficulty of the modeling process. Therefore, the search for novel robust, and reliable models is always the interest of hydrologists as such models will ensure better river engineering and sustainability.

The main aim of this research is to provide state of the art multi-strategy hybrid expert system consisting of the EWT decomposition technique with five FS approaches (i.e., BRF, SA, RR, relief algorithm, and mutual information) and two advanced ML methods (EDBi-LSTM and XGB) to forecast the daily SSL (during 07/01/2001–31/10/2015) in two study sites of Godavari river basin, India. Unlike the previous literature, which focuses on the individual enhancing strategy in the hybrid forecasting ML-based models, we investigated several powerful FS approaches to evaluate the accuracy enhancement and computational cost reduction compared to the hybrid models with no FS and standalone counterpart models (EDBi-LSTM and XGB). The most significant antecedent time-lagged inputs applied in the EWT decomposition procedure were extracted by the PACF method. The effectively filtered sub-sequences by the FS approaches were used to feed the ML models. The selected case study in the Godavari river concerning the high signal fluctuations related to sediment data in the monsoon seasons can appropriately prove the actual robustness of the developed forecasting hybrid models.

Table 1

Geographical coordinates of the two understudy stations in the Godavari river basin (GRB) and descriptive statistics of daily suspended sediment load during (2010–2015).

Ashti	Latitude (N) = 19° 41′ 04′′, Longitude (E) = 79° 47′ 10′′, and Elevation (m) = 141.42			
Bamini	Latitude (N) = 19° 48′ 47′′, Longitude (E) = 79° 22′ 52′′, and Elevation (m) = 157.97			
Metric	Ashti		Bamini	
	Training (2001–2010)	Testing (2011–2015)	Training (2001–2010)	Testing (2011–2015)
Minimum	0.0010	0.0150	0.000	0.020
Maximum	3.940	1.200	4.728	2.000
Mean	0.2536	0.1660	0.2378	0.2795
Std. deviation	0.2911	0.1632	0.4578	0.1645
Skewness	3.597	2.164	4.248	3.500
Kurtosis	27.62	6.338	23.67	29.72

2. Material and methods

2.1. Study area and statistically data description

Fig. 1 shows the location map of the Godavari river basin (GRB) along with the two study sites, i.e., Ashti and Bamini, selected for daily suspended sediment load (SSL) forecasting using the proposed hybrid expert system. The GRB (longitude 73° 24′ 00′′ to 83° 07′ 00′′ E, and latitude 16° 19′ to 22° 34′ 00′′) originates from Trimbakeshwar village of Nashik district of Maharashtra, India. It covers an area of 312812 km², approximately 9.5% of the total catchment area, with a length of 1465 km. The elevation of the basin ranges from 246 m to 1677 m and spreads over the states of Maharashtra (48.6%), Andhra Pradesh (23.4%), Madhya Pradesh (10.0 %), Chhattisgarh (10.9%), Orissa (5.7%),

Karnataka (1.4%), and Puducherry (0.001%). It flows from the southeast direction and falls into the Bay of Bengal. The average annual rainfall in the entire GRB varies from 755 mm to 1531 mm. Most (84%) of it is received during the southwest monsoon (June–September). The mean minimum and maximum air temperature fluctuate from 20.63 °C to 33.04 °C. The major tributaries of the GRB are Pranhita, Wainganga, Wardha, Penganga, Maner, Manjira, Pravara, Purna, Mula, Indravati, and Sabari. Table 1 provides information about the geographical coordinates of the study sites and the length of data.

The monsoon season daily SSL data for 15 years (July 1, 2001-October 31, 2015) were acquired from the India-WRIS (Water Resources Information System). The available data was partitioned into two sets: (i) From July 1, 2001, to October 31, 2010, for training, and (ii) For testing the duration from July 1, 2011 to October 31, 2015 is
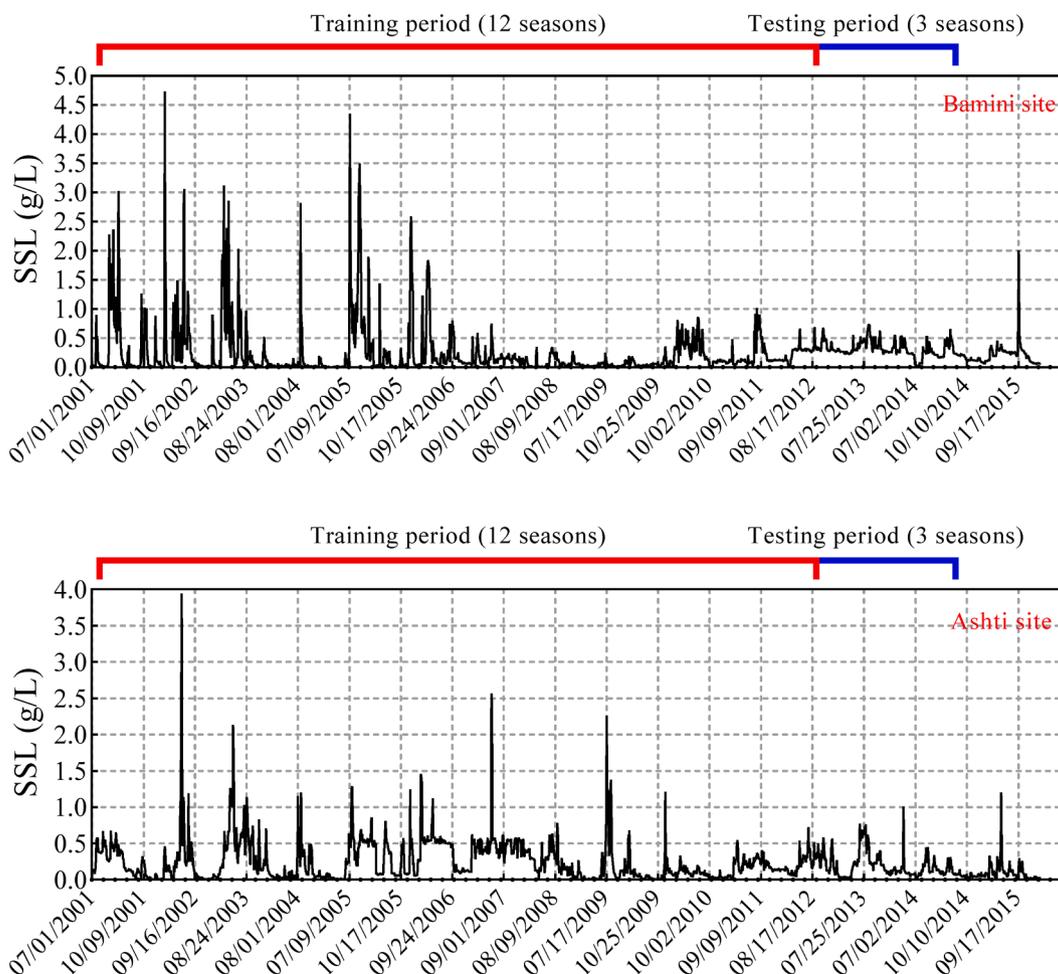
Fig. 2. Real time seasonal signals of SSL for Bamini and Ashti sites.

selected. The seasonal variation of SSL time-series data is illustrated in Fig. 2, while Table 1 summarizes the data's basic statistics during training and testing periods.

2.2. Empirical wavelet decomposition (EWT)

This paper used the Empirical Wavelet Transform (EWT) to decompose a time series of suspended sediment load. According to Jerone Gilles's (2013) and Hu and Wang's (2015) EWT approach is a revolutionary strategy for removing the stochastic volatility inherent in the original time series data (Gilles, 2013; Hu and Wang, 2015). Only the important components of the original suspended load series are retrieved during this operation. The EWT approach is based on resilient pre-processing of the peak frequency component followed by maxima-based spectrum segmentation, which constructs a matching wavelet filter band (Park et al., 2019). The EWT approach finds and extracts time series data's inherent modes. This procedure introduces a unique way of constructing a family of wavelets that represents and decomposes the processed data signal into a finite number of modes. The algorithm for decomposing signals with EWT is as follows (Liu et al., 2021; Tao et al., 2021b):

1- Determine the Fourier spectrum $F(w)$ of a coordinate series $\times$ by examining the frequency spectrum $x(t)$ with the deformation simulation.

2- The boundary of the Fourier spectrum $\Omega = \{\omega^n\}_{n=0}, 1, \cdots, N$ is determined by its segmentation.

One way to decompose a Fourier spectrum $[0, \pi]$ is to look at it in pairs of adjacent subsets (n). In this case, $\omega^0 = 0$ and $\omega^n = \pi$.

3- Following that, low-pass and band-pass filters are constructed. The empirical scaling function $O_n(\omega)$ and the empirical wavelet function $\psi_n(\omega)$ are denoted by eq. (1) and (2), respectively.

$$O_n(\omega) = \begin{cases} 1; & |\omega| \leq \omega_1 - \tau_1 \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_1}(|\omega| - \omega_1 + \tau_1)\right)\right]; & \omega_1 - \tau_1 \leq |\omega| \leq \omega_1 + \tau_1 \\ 0; & \text{others} \end{cases}$$
(1)

$$\psi_n(\omega) = \begin{cases} 1; & \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}(|\omega| - \omega_{n+1} + \tau_{n+1})\right)\right]; & \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_{n+1} + \tau_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n + \tau_n)\right)\right]; & \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0; & \text{others} \end{cases}$$
(2)

Here, ω is the frequency, ω_n is the nth border frequency, τ is the parameter that prevents two continuous transformations from overlapping, and $\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

4- Different modes are extracted using empirical scale and empirical wavelet functions, and the resultant scale coefficient is the inner product of the scale function and the observed data. The coefficient is depicted in Eq. (1). Additionally, the empirical wavelet coefficients are expressed as the inner product of the empirical wavelet function and the measured data:

$$W_f^\varepsilon(0, t) = \langle f, O_1 \rangle = \int f(\tau)\overline{O_1(\tau - t)}d\tau = F^{-1}(\omega)[f(\omega)\widehat{O}_1(\omega)]$$
(3)

$$W_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\tau)\overline{\psi_n(\tau - t)}d\tau = F^{-1}(\omega)[f(\omega)\widehat{\psi}_k(\omega)]$$
(4)

Here, τ is the time node; $O_1(t)$ and $\psi_n(t)$ are the empirical scaling and empirical wavelet functions, respectively; $\widehat{O}_1(\omega)$ and $\widehat{\psi}_k(\omega)$ are the Fourier transforms of $O_1(t)$ and $\psi_k(t)$, respectively; and $\overline{O}_1(\tau - t)$ and $\overline{\psi}_k(\tau - t)$ are the conjugate complex numbers of $O_1(\tau - t)$ and $\psi_k(\tau - t)$, respectively.

Finally, the phrase for reconstructing original data is as follows:

$$f(t) = W(0, t)*O_1(t) + \sum_{n+1}^{K} W(n, t)*\psi_k(t)$$
(5)

Convolution is indicated by the symbol *. As a result, the EWT decomposition has the following IMF (intrinsic mode function):

$$f_0(t) = W(0, t)*O_1(t)$$
(6)

$$f_1(t) = W(i, t)*\psi_i(t)$$
(7)

2.3. Boruta-random forest feature selection

The stage of optimal feature selection is critical for implementing ML algorithms in modeling nonlinear hydrological processes, and this issue has been tackled by the Boruta-random forest (BRF) algorithm for feature selection (Kursa et al., 2010). The BRF algorithm computes the Z-scores for each input predictor concentrating on the shadow property. The distribution of Z-score metrics exposes the basic features of the predictors (Kursa et al., 2010). Many successful applications of the BRF algorithm have been found in different domains (Ahmed et al., 2021; Jamei et al., 2022b; Leutner et al., 2012; Li et al., 2016; Lyu et al., 2017; Prasad et al., 2019). The development of the BRF algorithm involves the following major steps (Kursa and Rudnicki, 2010):

1. 1 To start, it randomizes the input data set by making scrambled (shadow) duplicates of all features.
2. Compute the mean decrease accuracy (MDA) for each feature trained using a random forest classifier on the larger data set. The MDA is calculated by the following eq. (Hur et al., 2017; Strobl et al., 2008):

$$MDA = \frac{1}{m_{tree}} \sum_{m=1}^{m_{tree}} \frac{\sum_{t\in OOB} I(y_t = f(x_t)) - \sum_{t\in OOB} I(y_t = f(x_t^n))}{|OOB|}$$
(8)

In which OOB represents out-of-bag (i.e., the prediction error for each of the training trials aggregated by bootstrap), whereas $(y_t = f(x_t))$ and $(y_t = f(x_t))$ signify the predicted values before and after permutation, separately. Additionally, $I(\bullet)$ designates the indicator function.
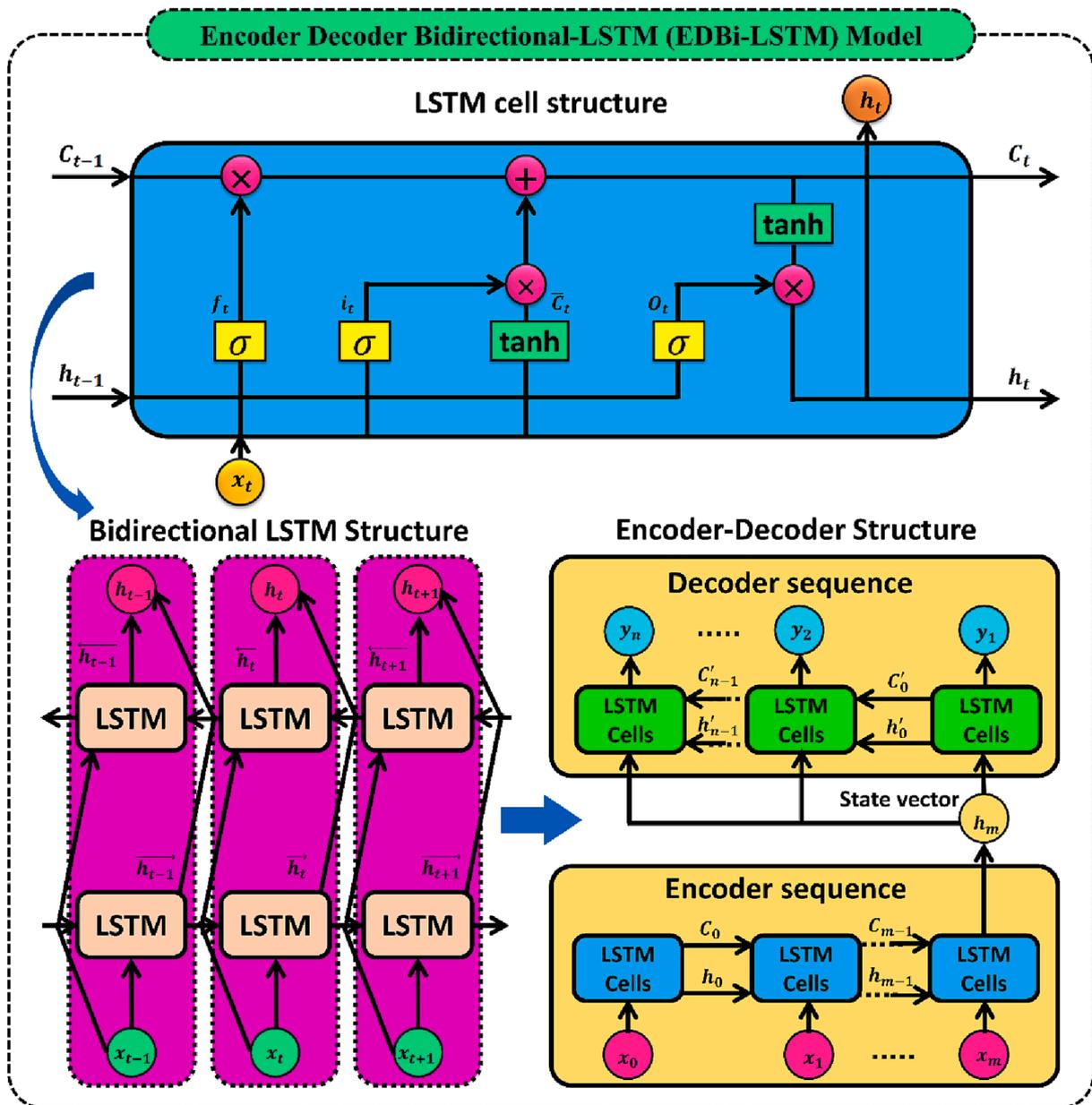
Fig. 3. The internal structure of the (a) LSTM, (b) Bi-LSTM, and (c) EDBi-LSTM models. Note: h_{t-1}, and C_{t-1} designates the hidden state output and network memory unit at the previous time step ($t-1$).

3. Calculate the Z-score for each iteration (i.e., determines a genuine feature is more essential than the best of its shadow features and repeatedly eliminate features supposed to be very irrelevant):

$$Z-score = \frac{MDA}{std} \tag{9}$$

Here, std = standard deviation of accuracy losses; after that, the maximum Z-score for duplicate attributes (MZSA) was computed.

4. If Z-scores < MZSA, the inputs are tagged "unimportant" and disconnected permanently until inputs with Z-scores > MZSA are designated "Confirmed.".
5. To end, the method terminates when all features have been validated or the required number of random forest iterations has been reached.

2.4. Relief feature selection

Kira and Rendell (1992) discovered the concept of the relief feature selection (RFS) approach for eliminating redundant and irrelevant features from the data set in simulating the classification and regression problems. This way reduced the computational cost, overfitting, complexity, and acceptable accuracy based on minimum features. Based on the relationship among the created model, the RFS methods are of three types (i) filter, (ii) wrapper, and (iii) embedded (Urbanowicz et al., 2018). It is an iterative non-deterministic procedure that randomly picks a subset of training samples in the feature space to capture the degree of relevancy among the elected features and target. The fundamental concept of the RFS method is to govern the important weight of each nominated feature in the sampled instance to evaluate the unique ability among the class labels. Additionally, the more positive weights, the more predictive attributes (Malik and Yadav, 2021). The application of the RFS involves the following steps:

Step-1: Start all feature weights ($w_i = 0$).

Step-2: Iteratively select a random set of instances from the training data or in each instance sample and assume the number of nearest neighbor samples (k).

Step-3: Calculate the Euclidean distance between desired feature and target using the k-nearest neighbor for each class and display the allocated maximum weight.

Step-4: Iteratively updates the weights of ith attributes in such a way as to find the nearest Hit (identical class) and nearest Miss (opposite class) instances for the examined samples as (Amjady and Keynia, 2009):

$$w_i = w_i + |x^i - nearestMiss^i x| - |x^i - nearestHit^i x|, i = 1, 2, \cdots, I \qquad (10)$$

where, $x^i = i^{th}$ attributes of the selected samples x, and $I =$ the number of candidate input features.

Step-5: Standardize the weight score values, average the updated weights in all iterations, and attain the final weight value. As a rule of thumb, the value of k can be obtained based on the number of data under training and obtained using the following expression (Amjady and Keynia, 2009):

$$k = Round(log_2(N)) \qquad (11)$$

where, $Round(.)$ denotes a function that rounds the real number to the closest integer value, and $N =$ the number of samples or training patterns.

2.5. Simulated annealing (SA) feature selection

The simulated annealing (SA) algorithm for feature selection is a non-deterministic population-based approach used to find a better solution for an optimization problem (Jeong et al., 2016). Over time, this algorithm received many applications in several fields (Ali et al., 2021; Debuse and Rayward-Smith, 1997; Meiri and Zahavi, 2006). The execution of SA includes the following steps:

1. Create an appropriate random solution in the primary state.
2. Calculate the cost in step (1) by employing the cost function.
3. Further, calculate a random neighboring solution.
4. Compute the cost for the above solution in step (3).
5. If the $cost(a) > cost(b)$, produce a new solution; otherwise, go to step (6).
6. Repeat steps (3–5) until an optimal solution is exposed.

It should be mentioned that the main tuning parameters for this type of feature selection are population, maximum iteration, initial temperature, temperature reduction rate, and desired feature number.

2.6. Ridge regression (RR) feature selection

The ridge regression (RR) method of feature selection was introduced by Hoerl and Kennard (1970) to handle collinearity problems, which may happen in moderate, exact, complete, and severe collinearity amongst highly correlated regressors or predictors. Considering a set of k features with N observtaions, each is standardized with zero mean and variance of one. The cost function for RR is employed to estimate the robust features using the following expression (Jamei et al., 2022a):

$$min_{(\omega_0, \omega)} \left[\frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \omega_0 - x_i^T \omega \right)^2 + \frac{\lambda}{N} R(\omega) \right] \qquad (12)$$

In which, ω represents ($\omega_1, \omega_2, \ldots, \omega_\kappa$) model parameters; $R(\omega)$ denotes a penalty term regulated by the λ; while $\sum_{i=1}^{N} \left(y_i - \omega_0 - x_i^T \omega \right)^2$ designates the square loss term. The ridge-regression penalty term (aka L_2 norm) is written as(Shahsavar et al., 2021):

$$R(\omega) = \frac{1}{2} \|\omega\|_2^2 = \frac{1}{2} \sum_{\rho=1}^{\kappa} \omega^{\rho^2} \qquad (13)$$

The penalty term of the RR shrinks the regression coefficients size whereby a multidimensional sphere characterizes the constraint region; hence, the elliptical contours can intersect anywhere on the sphere. As a successor algorithm of the ordinary least squares regression, the RR disables the collinearity problems by penalizing the least-squares loss on the regression coefficients by an L_2 penalty (Wu, 2020).

2.7. Mutual information

Mutual information is a measure of information that is used to quantify the shared information between two random variables. It may also be used to determine the correlation between two random variables. Mutual information is derived from the idea of information entropy. The information entropy of random variables measures their uncertainty (Kwak and Choi, 2002). The random variable X's entropy is given as (Vergara and Estévez, 2014):

$$S(X) = - \sum_{x \in X} P(x) log P(x) \qquad (14)$$

Here, $P(x)$ denotes the probability of the variable. The entropy of variables is only connected to their distribution, not to the variables themselves. As a result, interference from noise can be avoided. The mutual information between two random variables X and Y is represented by their joint probability distribution, which is denoted by (Tian et al., 2019):

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) log \frac{P(x,y)}{P(x)P(y)} \qquad (15)$$

Where $P(x,y)$ denotes the joint probability distribution of random variables X and Y, and $P(x)$ and $P(y)$ respectively represent the marginal probability distributions of random variables X and Y.

2.8. Encoder decoder Bidirectional- LSTM (EDBi-LSTM)

In the deep learning model, the LSTM (long short-term memory) technique recently became popular in different domains (Bhattarai et al., 2023; Chen et al., 2020; Feng et al., 2020; Mengel, 1993; Xiang et al., 2020; Yin et al., 2020). It is based on the architecture of the recurrent neural network (RNN), and the extension to that model is called bidirectional-LSTM (Hochreiter and Schmidhuber, 1997). In LSTM, the input data is processed based on previous information (Kim and Kim, 2020). The LSTM network comprises a forgetting gate, input gate, output gate, hidden state, and cell state (Fig. 3). In each step, LSTM updates the six parameters which are described (Elsayed et al., 2023; Livieris et al., 2020):

$$f_t = \sigma\left(w_f \bullet h_{t-1} x_t + b_f\right) \qquad (16)$$

$$i_t = \sigma\left(w_i \bullet h_{t-1} x_t + b_i\right) \qquad (17)$$

$$\overline{C}_t = tanh\left(w_c \bullet h_{t-1} x_t + b_c\right) \qquad (18)$$

$$C_t = f_t \times C_{t-1} + i_t \times \overline{C}_t \qquad (19)$$

$$O_t = softsign\left(w_o \bullet h_{t-1} x_t + b_0\right) \qquad (20)$$

$$h_t = O_t \times RLU(C_t) \qquad (21)$$

here, f_t, i_t, $\overline{C}_t$, C_t, O_t, and $h_t =$ forget gate, input gate, new cell state candidate vectors, cell state, output gate, and final output. $x_t =$ input at time t, $\sigma =$ sigmoid function of f_t and i_t. Furthermore, w_f, w_i, w_c, w_o, and b_f, b_i, b_c, and b_0 are weights matrix and bias vectors of f_t, i_t, $\overline{C}_t$, and O_t,
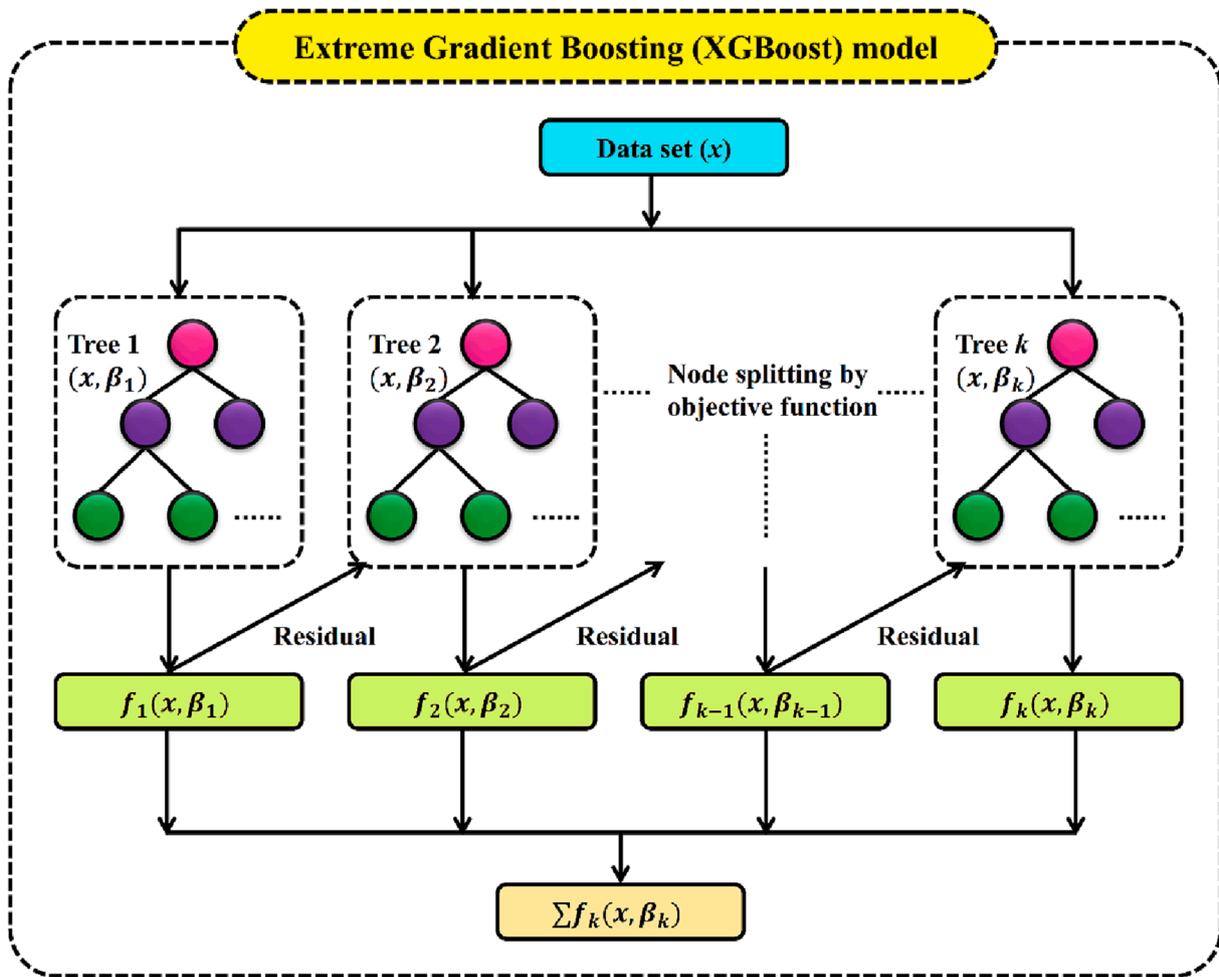
Fig. 4. Schematic flowchart of XGB model.

separately. The *tanh* (hyperbolic: −1, 1), *softsign*, and ReLU (Rectified Linear Unit) are activation function of $\overline{C}_t$, O_t, and h_t units.

In LSTM, when backpropagation is carried out, the gradient explosion and information morphing drawbacks are exposed (Yin et al., 2020). All these issues were tackled by the Bi-LSTM model and effectively processed the sequential data composed of time instances (Graves and Schmidhuber, 2005; Hu and Zhang, 2018). However, in many cases, it is necessary to know the impact of SSL on previous time steps other than the SSL at the current time step of prediction. Thus, this study utilized the encoder-decoder bidirectional-LSTM (EDBi-LSTM) model for predicting daily SSL on Ashti and Bamini sites. The concept of EDBi-LSTM was proposed by Cho et al. (2014) to recognize models on different input and output time stages. Fig. 3 illustrates the typical structure of the EDBi-LSTM model, which is composed of two distinct parts (i) one part for reading the input information of the sequence, and encoding the fixed-length vector, and (ii) a second part for decoding the vector and outputting the forecasted sequence. The final vector generated from the encoded LSTM with *m* time steps is stored in the cell state and acts as input for the decoder LSTM with *n* time steps. Thus, it can be said that during each update, the decoder feeds the model so that the output from the previous update is used as the input of the current update because this sequence-to-sequence network solves the time-step issue (Bian et al., 2020; Xiang et al., 2020; Zhang et al., 2021).

2.9. Extreme learning gradient boosting (XGB)

The XGB (extreme gradient boosting machine) is a supervised learning algorithm used to optimize the regression and classification

problems introduced by Chen and Guestrin (2016). It is the extended form of the GBT (gradient boosting tree) algorithm (Abba et al., 2020; Tao et al., 2022), and has become popular in many fields (Feigl et al., 2021; Ferreira and da Cunha, 2020; Sikorska-Senoner and Quilty, 2021). Trees in the XGB algorithm are created by implementing the three major steps: (i) the whole training data subgroup is fitted with a decision tree, (*ii*) the residuals of the model tree are fitted with a loss function, and (*iii*) add the loss function and model tree for generating the next tree (Brédy et al., 2020).

Additionally, the XGB algorithm is less prone to overfitting due to Taylor expansion and a regularization term (Guo et al., 2020). Fig. 4 illustrates the typical network of the XGB algorithm (*x* represents input vectors, and β_k denotes independent and identically distributed random vectors). The output of the XGB is written as (Chen and Guestrin, 2016; Ni et al., 2020):

$$\widehat{y} = \sum_{k=1}^{k} f_k(x_i), f_k \in F \tag{22}$$

In which, f_k denotes separate regression tree, *k* represents many regression functions, and *F* signifies regression tree space. Similarly, x_i and $\widehat{y}$ indicates input and predicted values. Furthermore, the minimized regularized objective to learn the set of functions used in the model is expressed as:

$$\delta = \sum_i l(\widehat{y}_i, y_i) + \sum_k \mu(f_k) \tag{23}$$

Here, *l* labels differentiable convex loss function, μ designates

Table 2

Tuning the hyperparameters of ML models and outcomes of each FS strategy for the Ashti site.

Model	ML (E_{DBi}-LSTM and XGB) setting	Pre-processing setting
E_{WT}-E_{DBi}-LSTM	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.0008 Training Algorithm: Adam, Epochs: 80 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 27 Reduction percentage: 0%
E_{WT}-E_{DBi}-LSTM-Boruta	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.01 Training Algorithm: Adam, Epochs: 70 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 17 Reduction percentage: 37%
E_{WT}-E_{DBi}-LSTM-MI	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.005 Training Algorithm: Adam, Epochs: 80 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 14 Reduction percentage: 48%
E_{WT}-E_{DBi}-LSTM-RR	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.007 Training Algorithm: Adam, Epochs: 70 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 13 Reduction percentage: 52%
E_{WT}-E_{DBi}-LSTM-SA	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.007 Training Algorithm: Adam, Epochs: 80 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 15 Reduction percentage: 44%
E_{WT}-E_{DBi}-LSTM-R_{elief}	Number of Layers: 2, Neurons number: 250, Learning Rate: 0.00075 Training Algorithm: Adam, Epochs: 80 Dropout: 0, activation='relu', Batch Size: 16	Sub-sequences number: 15 Reduction percentage: 44%
E_{DBi}-LSTM	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.0003 Training Algorithm: Adam, Epochs: 90 Dropout: 0, activation='relu', Batch Size: 16	Sub-sequences number: 3 Reduction percentage: 0%
E_{WT}-X_{GB}	Learning rate: 0.025, N_Estimatiors: 200, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 27 Reduction percentage: 0%
E_{WT}-X_{GB}-Boruta	Learning rate: 0.1, N_Estimatiors: 250, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 17 Reduction percentage: 37%
E_{WT}-X_{GB}-MI	Learning rate: 0.04, N_Estimatiors: 250, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 14 Reduction percentage: 48%
E_{WT}-X_{GB}-RR	Learning rate: 0.050, N_Estimatiors: 250, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 13 Reduction percentage: 52%
E_{WT}-X_{GB}-SA	Learning rate: 0.060, N_Estimatiors: 300, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 15 Reduction percentage: 44%
E_{WT}-X_{GB}-R_{elief}	Learning rate: 0.20, N_Estimatiors: 200, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 15 Reduction percentage: 44%
X_{GB}	Learning rate: 0.050, N_Estimatiors: 60, Max-Depth:7, min_child_weight = 4	Sub-sequences number: 3 Reduction percentage: 0%

regularization term and is calculated as $\mu(f) = \gamma T + 1/2\lambda\|\omega\|^2$, where γ is each leaf complexity, T is the number of leaves in a decision tree, λ denotes the regularization hyper-parameter, and ω is the leaves score vector. Moreover, y_i represents observed value, $\hat{y}_i$ indicates predicted value, and δ is regularized objective function. Tables 2 and 3 outline the optimized values of all the hyperparameters (i.e., the maximum tree depth, learning rate, minimum sum of instance weight needed in child, and the number of estimators) of the XGB model used daily SSL prediction at Ashti and Bamini sites.

2.10. Goodness-of-fit metrics

Five statistical indices including correlation coefficient (R), root mean square error (RMSE), relative absolute error (RAE), Kling-Gupta

efficiency (KGE) (Gupta et al., 2009; Jamei et al., 2023b), Nash–Sutcliffe efficiency (NSE) (McCuen et al., 2006; Nash and Sutcliffe, 1970), and uncertainty coefficient with 95% confidence level ($U_{95\%}$) (Patino and Ferreira, 2015) were utilized to assess the efficacy and accuracy of the models for SSL forecasting:

$$R = \frac{\sum_{i=1}^{N} \left(SSL_{o,i} - \overline{SSL_o}\right)\left(SSL_{f,i} - \overline{SSL_f}\right)}{\sqrt{\sum_{i=1}^{N} \left(SSL_{o,i} - \overline{SSL_o}\right)^2 \sum_{i=1}^{N} \left(SSL_{f,i} - \overline{SSL_f}\right)^2}} \quad (24)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(SSL_{o,i} - SSL_{f,i}\right)^2} \quad (25)$$

Table 3
Tuning the hyperparameters of ML models and outcomes related to each FS strategy for the Bamini site.

Model	ML (E_{DBi}-LSTM and XGB) setting	Pre-processing setting
E_{WT}-E_{DBi}-LSTM	Number of Layers: 3, Neurons number: 250, Learning Rate: 0.005 Training Algorithm: Adam, Epochs: 90 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 36 Reduction percentage: 0%
E_{WT}-E_{DBi}-LSTM-Boruta	Number of Layers: 2, Neurons number: 250, Learning Rate: 0.0003 Training Algorithm: Adam, Epochs: 90 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 12 Reduction percentage: 67.5%
E_{WT}-E_{DBi}-LSTM-MI	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.005 Training Algorithm: Adam, Epochs: 90 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 12 Reduction percentage: 67.5%
E_{WT}-E_{DBi}-LSTM-RR	Number of Layers: 2, Neurons number: 250, Learning Rate: 0.001 Training Algorithm: Adam, Epochs: 90 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 15 Reduction percentage: 58%
E_{WT}-E_{DBi}-LSTM-SA	Number of Layers: 2, Neurons number: 250, Learning Rate: 0.006 Training Algorithm: Adam, Epochs: 90 Dropout: 0, activation='relu', Batch Size: 128	Sub-sequences number: 18 Reduction percentage: 50%
E_{WT}-E_{DBi}-LSTM-R_{elief}	Number of Layers: 3, Neurons number: 250, Learning Rate: 0.0005 Training Algorithm: Adam, Epochs: 80 Dropout: 0, activation='relu', Batch Size: 16	Sub-sequences number: 27 Reduction percentage: 25%
E_{DBi}-LSTM	Number of Layers: 2, Neurons number: 200, Learning Rate: 0.0005 Training Algorithm: Adam, Epochs: 80 Dropout: 0, activation='relu', Batch Size: 16	Sub-sequences number: 3 Reduction percentage: 0%
E_{WT}-X_{GB}	Learning rate: 0.03, N_Estimatiors: 150, Max-Depth:5, min_child_weight = 3	Sub-sequences number: 36 Reduction percentage: 0%
E_{WT}-X_{GB}-Boruta	Learning rate: 0.1, N_Estimatiors:150, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 12 Reduction percentage: 67.5%
E_{WT}-X_{GB}-MI	Learning rate: 0.1, N_Estimatiors: 250, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 12 Reduction percentage: 67.5%
E_{WT}-X_{GB}-RR	Learning rate: 0.080, N_Estimatiors: 250, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 15 Reduction percentage: 58%
E_{WT}-X_{GB}-SA	Learning rate: 0.060, N_Estimatiors: 300, Max-Depth:5, min_child_weight = 4	Sub-sequences number: 18 Reduction percentage: 50%
E_{WT}-X_{GB}-R_{elief}	Learning rate: 0.12, N_Estimatiors: 250, Max-Depth:6, min_child_weight = 4	Sub-sequences number: 27 Reduction percentage: 25%
X_{GB}	Learning rate: 0.050, N_Estimatiors: 70, Max-Depth:7, min_child_weight = 4	Sub-sequences number: 3 Reduction percentage: 0%

$$RAE = \frac{\left[\sum_{i=1}^{N}\left(SSL_{f,i} - SSL_{o,i}\right)^2\right]^{1/2}}{\left[\sum_{i=1}^{N}\left(SSL_{o,i}\right)^2\right]^{1/2}} \quad (26)$$

$$KGE = 1 - \sqrt{(R-1)^2 + \left(StD_f/StD_o - 1\right)^2 + \left(\overline{SSL_f}/\overline{SSL_o} - 1\right)^2} \quad (27)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(SSL_{o,i} - SSL_{f,i}\right)^2}{\sum_{i=1}^{N}\left(SSL_{o,i} - \overline{SSL_o}\right)^2} \quad (28)$$

$$U_{95\%} = 1.96\sqrt{SDe^2 + RMSE^2} \quad (29)$$

where $SSL_{f,i}$ denotes forecasted suspended sediment load and $SSL_{o,i}$ denotes actual suspended sediment load. $\overline{SSL_f}$ is the forecasted result's average and SDe denotes the standard deviation of error values. $\overline{SSL_o}$ is the average of the observed values of the suspended load. N signifies the total number of total samples. The standard deviation of observed and forecasted values are denoted by StD_o and StD_f, respectively. A perfect

model's Nash–Sutcliffe Efficiency (NSE) is one (NSE = 1). NSE = 0 implies that the model's forecasting performance is identical to the time series mean. When the observed mean is greater than the forecast of the model, the efficiency is less than zero (NSE < 0) (McCuen et al., 2006). The model performs well when the RMSE value is near zero and R, KGE, and IA values are close to unity.

3. Model development and configuration

Several researchers have been devoted to predict the SSL in rivers using metrological and hydrometric parameters in recent years. There is limited research on forecasting using the actual time series of SSL during times based on decomposition oriented ML models. Here a new complementary hybrid expert system comprised of EWT decomposition coupled with five feature selection methods (BRF, SA, RR, relief, and MI) and the state of the art ML approaches (EDBi-LSTM and XGB) to forecast daily SSL at two-zones during monsoon season (07/01/2001–31/10/2015) in Godavari river basin, India. In this research work, 12 hybridized models namely E_{WT}-E_{DBi}-LSTM, E_{WT}-X_{GB}, E_{WT}-E_{DBi}-LSTM-Boruta,
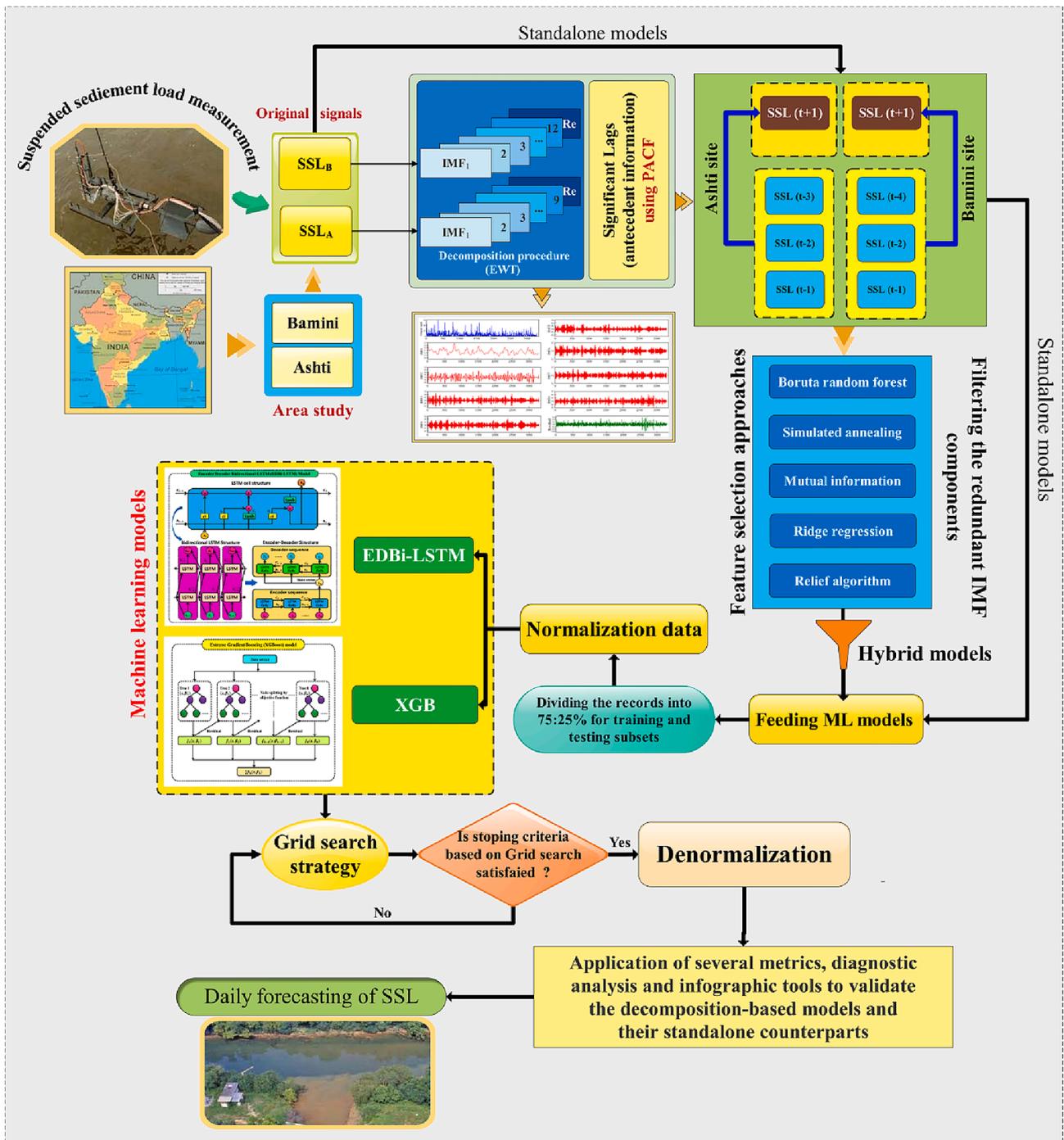
Fig. 5. Workflow of forecasting the daily SSL using the multi-strategy decomposition-based expert systems and standalone counterpart.
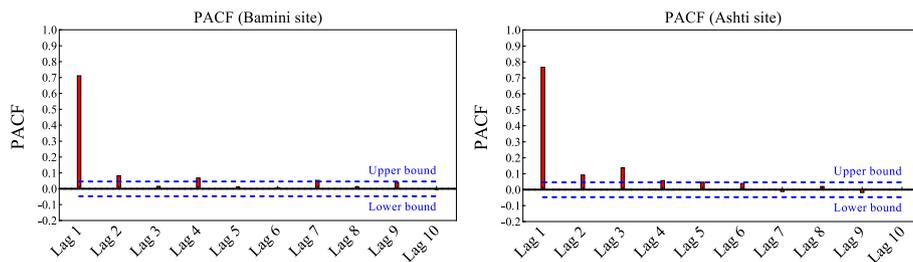
Fig. 6. Partial auto-correlation function of SSL signals for both Bamini and Ashi sites.
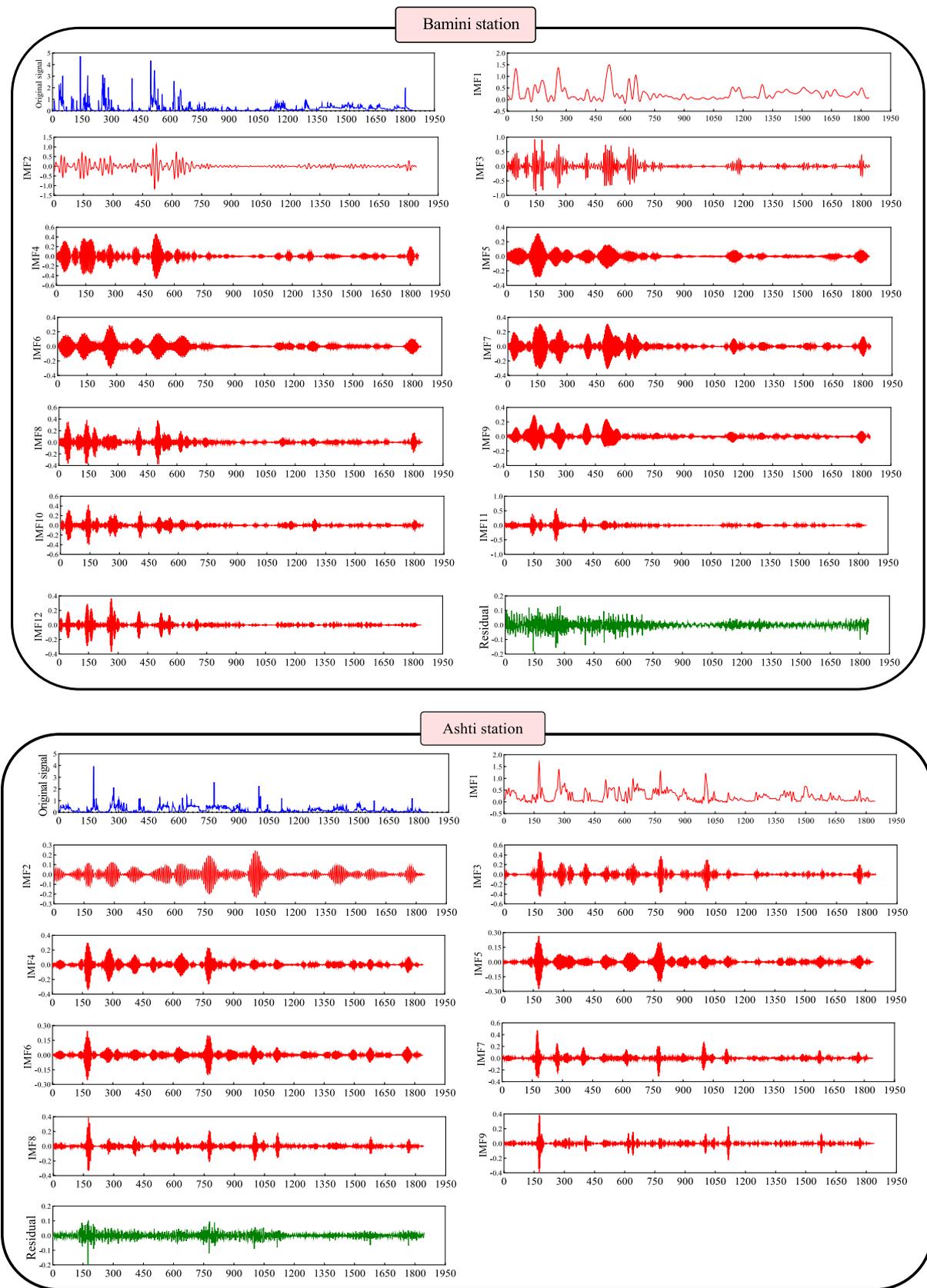
Fig. 7. Signal decomposition using the EWT technique into sub-signals for both sites.
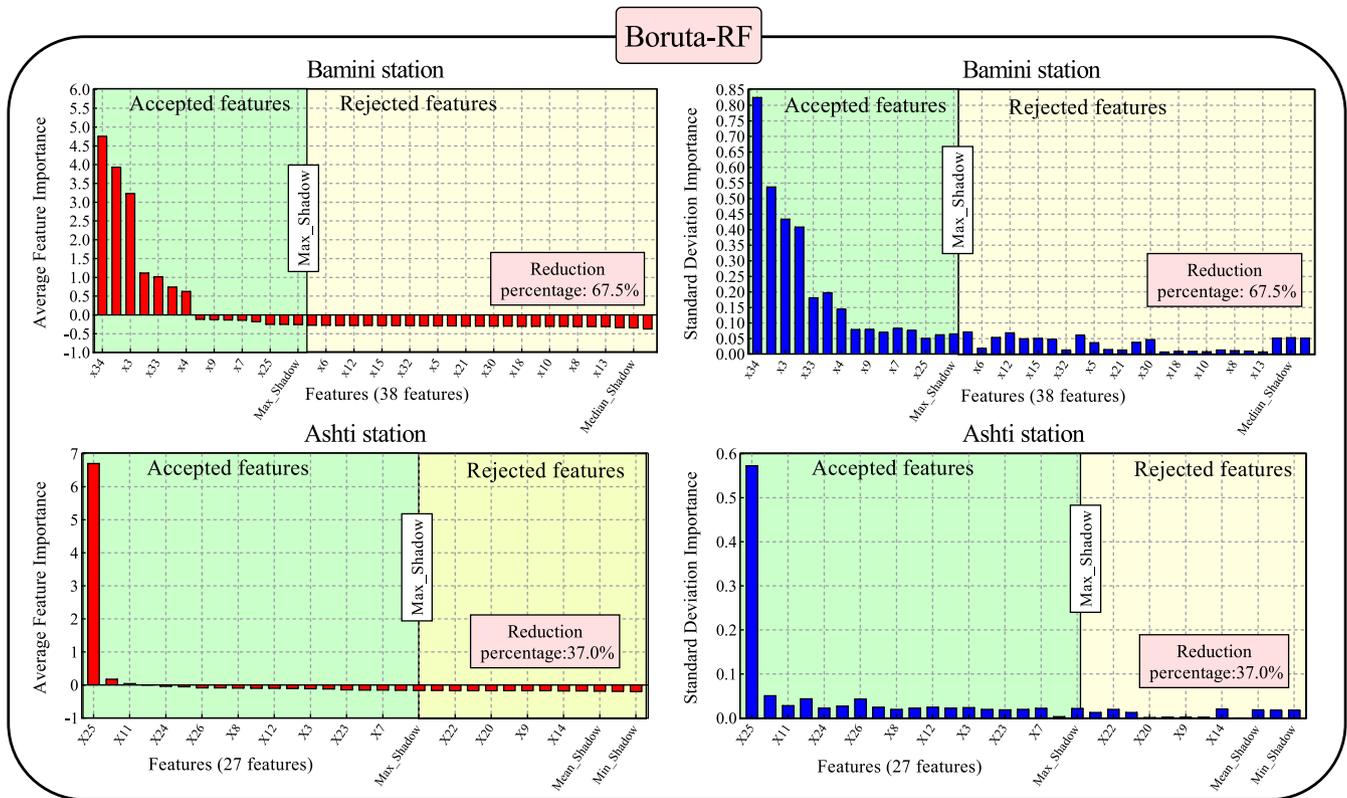
Fig. 8. Boruta- random forest outcomes to exploring the significant decomposed components (IMFs) to forecast the daily SSL in seasonal rivers of Bamini and Ashti sites.

E_{WT}-X_{GB}-Boruta, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-X_{GB}-SA, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-X_{GB}-RR, E_{WT}-E_{DBi}-LSTM-Relief, E_{WT}-X_{GB}-Relief, E_{WT}-E_{DBi}-LSTM-MI, and E_{WT}-X_{GB}-MI were constructed to forecast daily SSL in Bamini, and Ashti stations, India. All models were produced in MATLAB R2019a and Python environments. It is noteworthy that the XGB and EDBi-LSTM models were designed in Python 3.8 using the XGB and Tensorflow open-source software libraries.

Besides, the RR, MI, and BRF approaches were created in the Python platform based on the Scikit-learn (Pedregosa et al., 2011) and Boruta open-source libraries. In contrast, the relief and SA approaches were constructed in the Matlab environment. Execution of all the models was carried out by Intel (R) Core (TM) i7-6700 CPU @ 3.20 GHz computer system. Fig. 5 shows the workflow of the proposed hybrid approaches to forecast daily SSL in the two sites. The following steps of hybrid models are described in detail:

Step 1: Significant antecedent information gaining.

Since the only antecedent information associated with previous days of suspended sediment in both sites is used to forecast daily SSL, it is necessary to identify significant time delays. In this study, the PACF technique (Deo et al., 2016; Malik et al., 2019; Tiwari and Adamowski, 2013) is used to identify the significant lags which are reflected in Fig. 6. For the Bamini and Ashti sites, three antecedent time-lagged inputs (SSL (t-1), SSL(t-2), and SSL(t-4)) and (SSL(t-1), SSL(t-2), and SSL(t-3)) were identified as the most significant previous days information to design the hybrid models.

Step 2: Signals decomposition pre-processing.

This step is the most critical preprocessing procedure for the main signals using the EWT technique. In this step, first, the original time series is automatically decomposed into the sub-sequences (IMFs + residual). The number of IMFs depends on the time series characteristics. Then the antecedent information (time-lagged values) obtained from previous step is applied to the following sub-sequences. For the EWT decomposition, the "Sig2" type of signal, "global trend

removal"="none", "polynomial interpolation degree"=6, and "maximum number of bands"=3 as the decomposition adjustment was employed, which are recommended in the literature (Hu et al., 2015; Jamei et al., 2023a; Liu et al., 2018). As shown in Fig. 7, the number of IMFs components for the Bamini and Ashti sites were 12 and 7 respectively. Besides, all the sub-sequences obtained before the filtering stage for the mentioned sites were computed as 36 and 27 respectively.

Step 3: Determining the influential sub-sequences of FS schemes.

Feature selection (FS) methods in predictive models based on AI are generally used to aim for accuracy enhancement and to reduce the cost of computing. Feature selection methods have been frequently used in predicting engineering phenomena due to their special mechanism and handling of the uncertainty in the datasets. The effectiveness of these methods is still under study (Acikgoz, 2022). This research employs five well-known non-linear FS approaches to assess their potential in filtering the redundant decomposed sub-sequences used in ML methods for prediction.

First, the Baruta random forest (BRF) method was used to designate the hybrid forecasting models. In this method, the feature importance of each sub-sequence was determined by the logic presented in the methodology section and compared with the max_shadow benchmark value. Then, the redundant sub-sequences with less feature importance than the max_shadow benchmark value were excluded. Fig. 8 presents the results of the Boruta method containing the mean value and standard deviation of feature importance in both sites. RR, relief, and MI techniques are examined to filter the candidate inputs based on the importance factor criterion. In this regard, the sub-sequences with a higher importance factor than the corresponding value of cumulative frequency of 50% could be allocated to feed the ML approaches. As the last FS, the SA approach was adopted due to its ability to prevent trapping locally, enhance the population's diversity, and quickly converge to the global optimum (Ali et al., 2021). In the SA method, all the sub-sequences are ranked based on the best fitness, and the pre-defined number of sub-
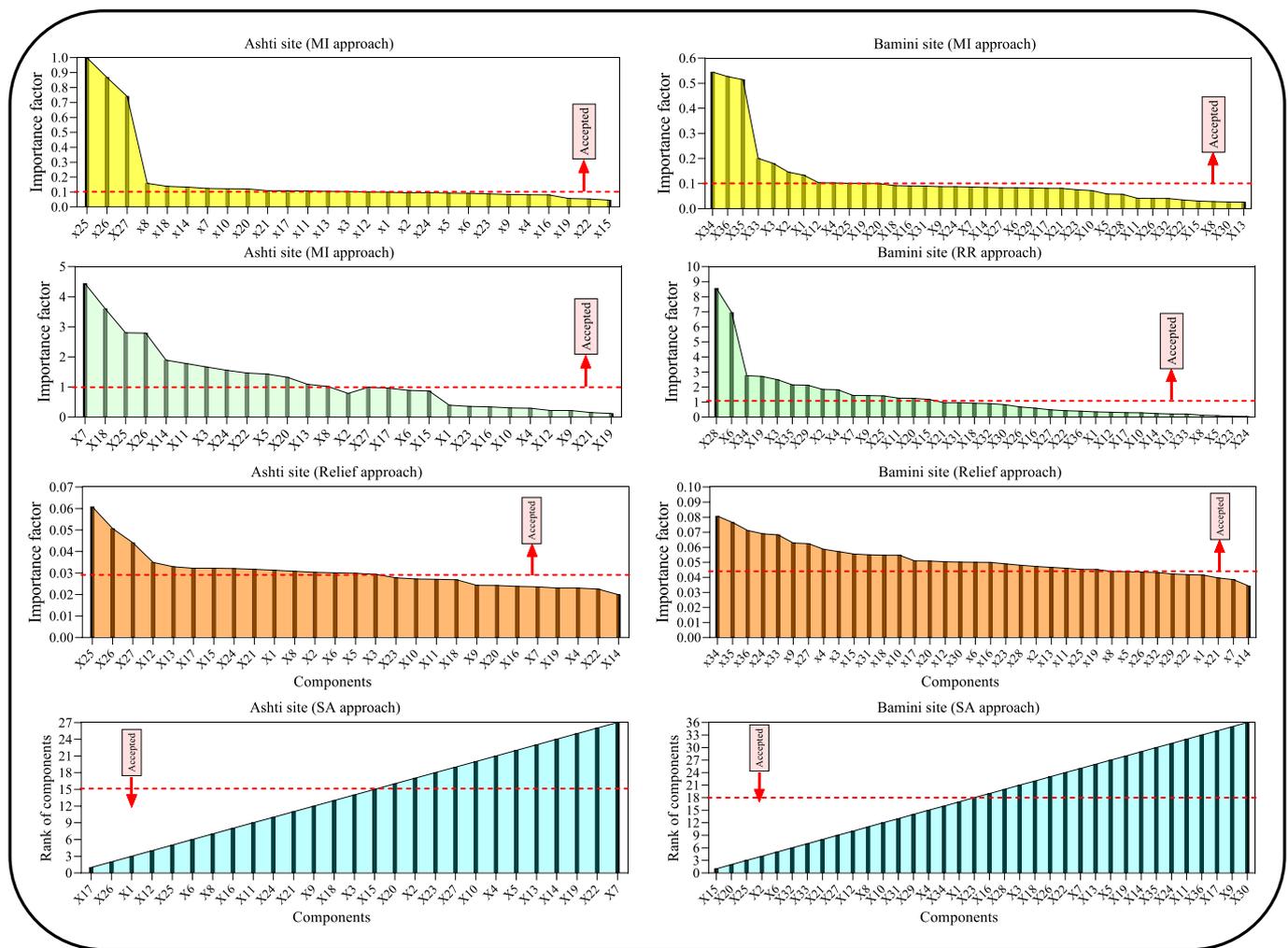
Fig. 9. Filtered decomposed sub-sequences using RR, MI, Relief, and SA feature selection techniques for Bamini and Ashti sites to feed the ML models.

sequences are selected to feed the ML models. For this purpose, we adjusted the following settings: Maximum iteration = 15, Population = 10, Initial temperature = 10, Temperature reduction rate = 0.99, and desired feature number = 15. Fig. 9 illustrates the outcomes of RR, relief, MI, and SA methods. According to the results of filtering the sub-sequences listed in Tables 2 and 3 in the site Ashti, the RR, MI, relief, SA, and BRF removed 52% and 48%, 44%, 44%, and 37% of the whole candidate sub-sequences respectively. In contrast in the Bamini site, the BRF, MI, RR, SA, and relief excluded 67.5% and 67.5%, 58%, 50%, and 20% of the sub-sequences respectively.

Step 4: Models feeding and configuration.

In this section, all the filtered decomposed components resulted from FS methods directly employed in the ML models. Here, 75% of whole datasets were allocated for training the models, and the rest of the datasets (25% of entire datasets) were used independently for the testing. Besides, normalization and denormalization procedures were implemented before feeding into the ML approaches to unite the data scale and enhance the convergence process in the training and testing stages (Ali et al., 2021). To examine the standalone EDBi-LSTM and XGB models, the second and third stages can be ignored, and only the antecedent time-lagged values of SSL for each site are employed to feed the ML models.

3.1. Setting of machine learning approaches

Optimal adjustment of ML models is very important in modeling engineering problems because the potential of different techniques can be adequately assessed (Rehamnia et al., 2020). Recently, there are common optimization approaches to gain the optimal setting hyperparameters of the model such as LSTM and XGB, which can be introduced as grid search (Shahsavar et al., 2021), random search, and trial integration with the metaheuristic optimization algorithms (Jamei et al., 2021). This research employed the grid search technique to optimize the implemented ML approaches to forecast the daily SSL in two selected sites. Previous literature shows that the most important setting hyperparameters for EDBi-LSTM (Bian et al., 2020; Zhang et al., 2021) and XGB (Singh et al., 2022; Zheng et al., 2017) are (Number of Layers, Neurons number in each layer, Learning Rate, Training Algorithm, Epochs value, Activation, and Batch Size) and (Max-Depth, Learning rate, and child_weight, N_Estimatiors) respectively. Tables 2 and 3 listed the optimal values of all the setting parameters of ML models in hybrid and standalone counterpart techniques in Bamini and Ashti respectively.

4. Application results and analysis

The forecasting of daily river suspended sediment load in the Bamini and Ashti sites, the performance accuracy of E_{DBi}-LSTM and its hybrid version, i.e., E_{WT}-E_{DBi}-LSTM, E_{WT}-E_{DBi}-LSTM-Boruta, E_{WT}-E_{DBi}-LSTM-

Table 4

Performance of the models for daily suspended sediment load forecasting in the Bamini site.

Model	Mode	R	RMSE	RAE	NSE	KGE	$U_{95\%}$
E_{WT}-E_{DBi}-LSTM	Training	0.9638	0.1188	0.2636	0.9286	0.9376	0.3293
	Testing	0.9212	0.0699	0.4039	0.8415	0.9029	0.1924
E_{WT}-E_{DBi}-LSTM-Boruta	Training	0.9951	0.0448	0.1027	0.9899	0.9753	0.1240
	Testing	0.9953	0.0171	0.1088	0.9905	0.9893	0.0473
E_{WT}-E_{DBi}-LSTM-MI	Training	0.9787	0.0915	0.2091	0.9577	0.9597	0.2537
	Testing	0.9283	0.0660	0.3836	0.8587	0.9057	0.1823
E_{WT}-E_{DBi}-LSTM-RR	Training	0.9642	0.1188	0.2569	0.9287	0.9496	0.3288
	Testing	0.9532	0.0546	0.3337	0.9034	0.9463	0.1514
E_{WT}-E_{DBi}-LSTM-SA	Training	0.9879	0.0696	0.1853	0.9755	0.9609	0.1926
	Testing	0.9586	0.0506	0.3334	0.9169	0.9572	0.1403
E_{WT}-E_{DBi}-LSTM-R_{elief}	Training	0.9579	0.1284	0.2878	0.9166	0.9197	0.3549
	Testing	0.9369	0.0635	0.3742	0.8692	0.9235	0.1747
E_{DBi}-LSTM	Training	0.7384	0.3000	0.4636	0.5449	0.6307	0.8315
	Testing	0.7707	0.1142	0.4252	0.5772	0.7460	0.3164
E_{WT}-X_{GB}	Training	0.9878	0.0806	0.2042	0.9672	0.8946	0.2233
	Testing	0.9372	0.0661	0.3984	0.8582	0.8315	0.1785
E_{WT}-X_{GB}-Boruta	Training	0.9954	0.0484	0.0773	0.9881	0.9425	0.1342
	Testing	0.9933	0.0217	0.1127	0.9847	0.9495	0.0601
E_{WT}-X_{GB}-MI	Training	0.9933	0.0526	0.1320	0.9860	0.9687	0.1458
	Testing	0.9156	0.0718	0.3938	0.8329	0.8439	0.1980
E_{WT}-X_{GB}-RR	Training	0.9922	0.0570	0.1467	0.9836	0.9613	0.1581
	Testing	0.9377	0.0619	0.3632	0.8757	0.8992	0.1705
E_{WT}-X_{GB}-SA	Training	0.9955	0.0430	0.1154	0.9906	0.9755	0.1193
	Testing	0.9541	0.0552	0.3509	0.9012	0.8962	0.1498
E_{WT}-X_{GB}-R_{elief}	Training	0.9970	0.0349	0.0965	0.9938	0.9843	0.0968
	Testing	0.9435	0.0593	0.3705	0.8861	0.9020	0.1630
X_{GB}	Training	0.8847	0.2231	0.4120	0.7482	0.6762	0.6183
	Testing	0.7540	0.1243	0.5250	0.4993	0.7501	0.3439
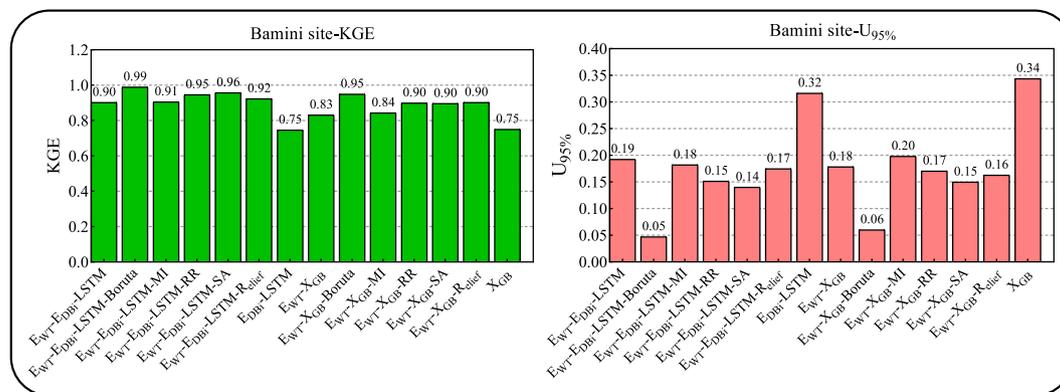
Fig. 10. Metric of KGE and U95% to assess the robustness of all the forecasting models in the Bamini site.

Boruta, E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-E_{DBi}-LSTM-R_{elief} and X_{GB} with its hybrid models E_{WT}-X_{GB}, E_{WT}-X_{GB}-Boruta, E_{WT}-X_{GB}-MI, E_{WT}-X_{GB}-RR, E_{WT}-X_{GB}-SA, and E_{WT}-X_{GB}-R_{elief} are examined for both the training and testing period to determine their forecasting capacity based on *R, RMSE, RAE, NSE, KGE,* and *$U_{95\%}$*.

4.1. Bamini site results analysis

This section offers a comprehensive assessment and evaluation in the Bamini site of the respective performances of the hybrid version as well as standalone models to forecast daily suspended sediment load. Table 4 describes that the hybrid E_{WT}-E_{DBi}-LSTM-Boruta model was out-performed against all other benchmarking comparing models to obtain the highest R, NSE, KGE, and lowest RMSE, RAE, and $U_{95\%}$ for the Bamini site to forecast daily suspended sediment load. The acquired magnitudes are E_{WT}-E_{DBi}-LSTM-Boruta [R = 0.9951, RMSE = 0.0448, RAE = 0.1027, NSE = 0.9899, KGE = 0.9753, $U_{95\%}$ = 0.1240]-train and [R = 0.9953, RMSE = 0.0171, RAE = 0.1088, NSE = 0.9905, KGE = 0.9893, $U_{95\%}$ =0.0473]-test period to forecast daily suspended sediment load. The second best model appeared to be E_{WT}-X_{GB}-Boruta, followed

by EWT-EDBi-LSTM-SA, E_{WT}-X_{GB}-SA, E_{WT}-E_{DBi}-LSTM-RR, and E_{WT}-X_{GB}-SA (see; Table 4). The E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-E_{DBi}-LSTM-R_{elief}, E_{WT}-E_{DBi}-LSTM, E_{WT}-X_{GB}-MI, E_{WT}-X_{GB}-R_{elief}, E_{WT}-X_{GB} models efficiency are reasonably good (see; Table 4) in both training and testing periods but could not exceed than the performance of the proposed E_{WT}-E_{DBi}-LSTM-Boruta model. Furthermore, the standalone E_{DBi}-LSTM and X_{GB} models accuracy is poor compared to the performance of their hybrid version of the models. Overall, the performance of the E_{WT}-E_{DBi}-LSTM-Boruta model confirmed its suitability and applicability in terms of daily sus-pended sediment load forecasting against the benchmarking comparing models for the Bamini site.

Fig. 10 shows the robustness of all the designed models for the Bamini site in terms of uncertainty variation (i.e., $U_{95\%}$) and KGE met-rics during the daily suspended sediment load forecasting. It is clearly visible that the E_{WT}-E_{DBi}-LSTM-Boruta model exhibited lower $U_{95\%}$ values while higher KGE values in the daily forecasting scenario as compared to E_{WT}-E_{DBi}-LSTM, E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-E_{DBi}-LSTM-R_{elief}, E_{WT}-E_{DBi}-LSTM, E_{DBi}-LSTM, E_{WT}-X_{GB}, E_{WT}-X_{GB}-Boruta, E_{WT}-X_{GB}-MI, E_{WT}-X_{GB}-RR, E_{WT}-X_{GB}-SA, E_{WT}-X_{GB}-R_{elief}, and X_{GB} models. The E_{WT}-E_{DBi}-LSTM-Boruta model for
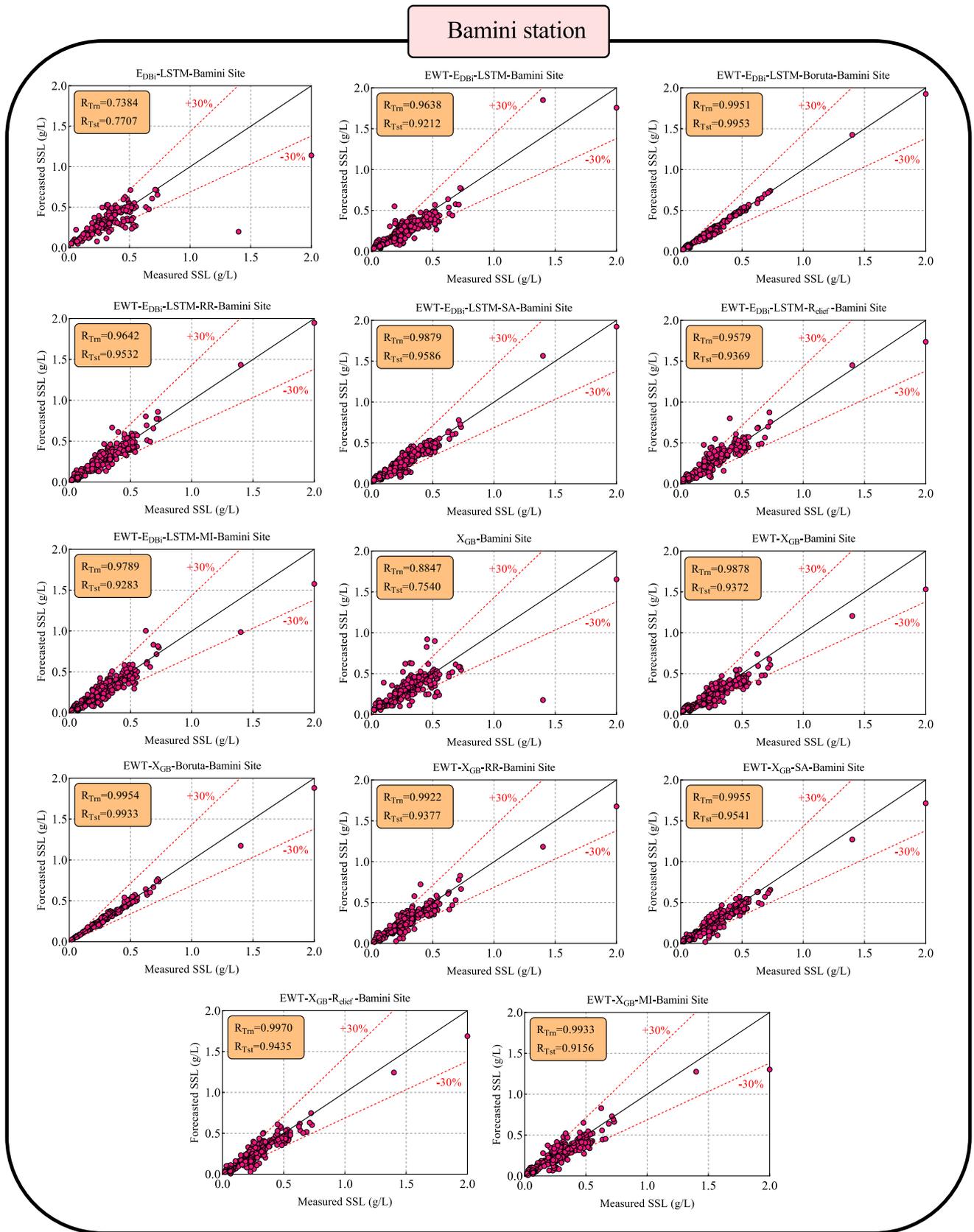
Fig. 11. Performance evaluating all the forecasting models in the form of scatter plots for the testing phase in Bamini site.

Table 5
Performance of the models for daily suspended sediment load forecasting in Ashti site.

Model	Mode	R	RMSE	RAE	NSE	KGE	$U_{95\%}$
E_{WT}-E_{DBi}-LSTM	Training	0.9643	0.0756	0.2453	0.9298	0.9473	0.2095
	Testing	0.9492	0.0505	0.3119	0.9001	0.9237	0.1397
E_{WT}-E_{DBi}-LSTM-Boruta	Training	0.9983	0.0176	0.0613	0.9962	0.9790	0.0478
	Testing	0.9981	0.0112	0.0801	0.9951	0.9649	0.0293
E_{WT}-E_{DBi}-LSTM-MI	Training	0.9910	0.0387	0.1246	0.9816	0.9657	0.1074
	Testing	0.9485	0.0509	0.3088	0.8983	0.8994	0.1412
E_{WT}-E_{DBi}-LSTM-RR	Training	0.9702	0.0691	0.2323	0.9413	0.9574	0.1916
	Testing	0.9560	0.0478	0.3026	0.9104	0.8879	0.1326
E_{WT}-E_{DBi}-LSTM-SA	Training	0.9812	0.0590	0.2022	0.9572	0.9290	0.1609
	Testing	0.9714	0.0390	0.2507	0.9402	0.9503	0.1077
E_{WT}-E_{DBi}-LSTM-R_{elief}	Training	0.9381	0.1004	0.3067	0.8761	0.8612	0.2783
	Testing	0.9124	0.0664	0.3609	0.8269	0.8199	0.1839
E_{DBi}-LSTM	Training	0.8095	0.1678	0.3467	0.6538	0.7277	0.4647
	Testing	0.7961	0.0994	0.4151	0.6126	0.7879	0.2756
E_{WT}-X_{GB}	Training	0.9812	0.0593	0.1730	0.9567	0.9017	0.1644
	Testing	0.9406	0.0553	0.3285	0.8800	0.8593	0.1534
E_{WT}-X_{GB}-Boruta	Training	0.9953	0.0280	0.0373	0.9904	0.9770	0.0776
	Testing	0.9967	0.0143	0.0833	0.9920	0.9653	0.0396
E_{WT}-X_{GB}-MI	Training	0.9839	0.0525	0.1652	0.9661	0.9371	0.1456
	Testing	0.9381	0.0560	0.3262	0.8770	0.8694	0.1552
E_{WT}-X_{GB}-RR	Training	0.9867	0.0475	0.1463	0.9722	0.9478	0.1318
	Testing	0.9480	0.0515	0.2918	0.8960	0.8855	0.1427
E_{WT}-X_{GB}-SA	Training	0.9855	0.0505	0.1494	0.9686	0.9332	0.1401
	Testing	0.9580	0.0465	0.2803	0.9152	0.9002	0.1288
E_{WT}-X_{GB}-R_{elief}	Training	0.9906	0.0398	0.1280	0.9805	0.9613	0.1103
	Testing	0.9477	0.0511	0.2961	0.8976	0.9076	0.1417
X_{GB}	Training	0.9137	0.1237	0.3353	0.8119	0.7476	0.3422
	Testing	0.7836	0.1026	0.5730	0.5867	0.6583	0.2800
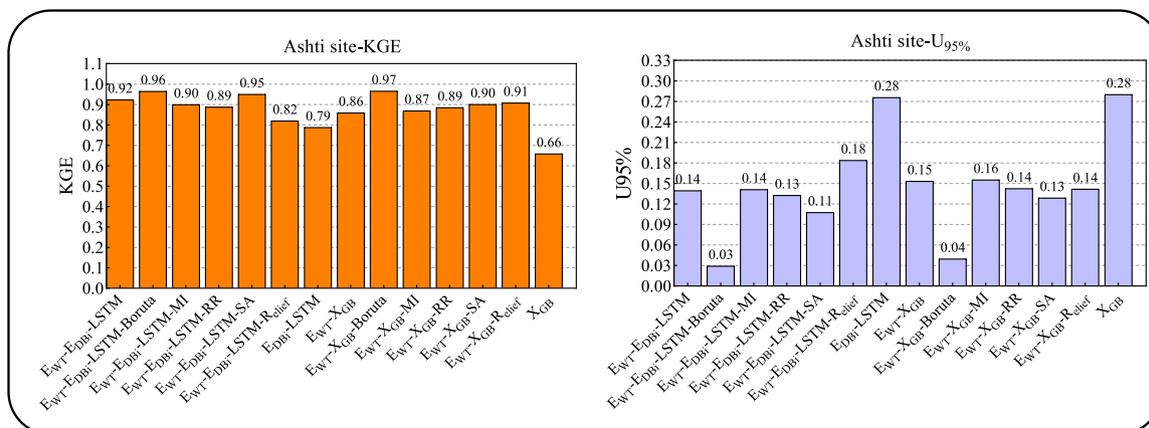
Fig. 12. KGE and U95% metrics to assess the robustness of all the forecasting models in the Ashti site.

example displays the values of KGE and $U_{95\%}$ = 0.99, 0.05%; followed by the E_{WT}-X_{GB}-Boruta model (2nd best) with KGE and $U_{95\%}$ = 0.95, 0.06% and so on. Again, the standalone models acquired relatively lower accuracy in terms of KGE and $U_{95\%}$ values as compared to hybrid models. Overall, Fig. 10 confirmed the superiority of the E_{WT}-E_{DBI}-LSTM-Boruta model against benchmark comparing models to forecast daily suspended sediment load for the Bamini site.

In the next Fig. 11, a scatter plot was drawn to show the better accuracy of the E_{WT}-E_{DBi}-LSTM-Boruta model over other comparable models between the daily measured and forecasted SSL. The scatter plot demonstrates the correlation between the forecasted and the measured suspended sediment load (i.e., SSL) using the coefficient of determination (R) in the training and testing period for the daily forecasting scenario in Bamini site. The E_{WT}-E_{DBi}-LSTM-Boruta model is giving reasonably good preciseness by generating a higher magnitude of R_{Trn} and R_{Tst} with respect to the benchmark hybrid comparing models E_{WT}-E_{DBi}-LSTM, E_{DBi}-LSTM, E_{WT}-E_{DBi}-LSTM-RR for daily suspended sediment lo forecasting. When comparing the E_{WT}-E_{DBi}-LSTM-Boruta model

against the standalone E_{DBi}-LSTM and X_{GB} models, Fig. 11 confirmed that the precision exhibited by the E_{WT}-E_{DBi}-LSTM-Boruta model is significantly better between th measured and forecasted SLL in terms of R_{Trn} and R_{Tst} Baminin site.

4.2. Ashti site results analysis

In this Table 5, the daily suspended sediment load forecasting for the Ashti site has been discussed and analyzed in detail. Here the precision of the E_{WT}-E_{DBi}-LSTM-Boruta and E_{WT}-X_{GB}-Boruta models are very close in both the training and testing period but the E_{WT}-E_{DBi}-LSTM-Boruta model is slightly better to achieve higher R, NSE, KGE and lower RMSE, RAE and $U_{95\%}$. For the training period, we refer to Table 5. While the acquired magnitudes of the hybrid version and the standalone models for testing periods are: E_{WT}-E_{DBi}-LSTM-Boruta [R = 0.9981, RMSE = 0.0112, RAE = 0.0801, NSE = 0.9951, KGE = 0.9649, $U_{95\%}$ = 0.0293] as compared to the E_{WT}-X_{GB}-Boruta, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-X_{GB}-SA, E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-X_{GB}-MI, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-X_{GB}-RR, E_{WT}-E_{DBi}-
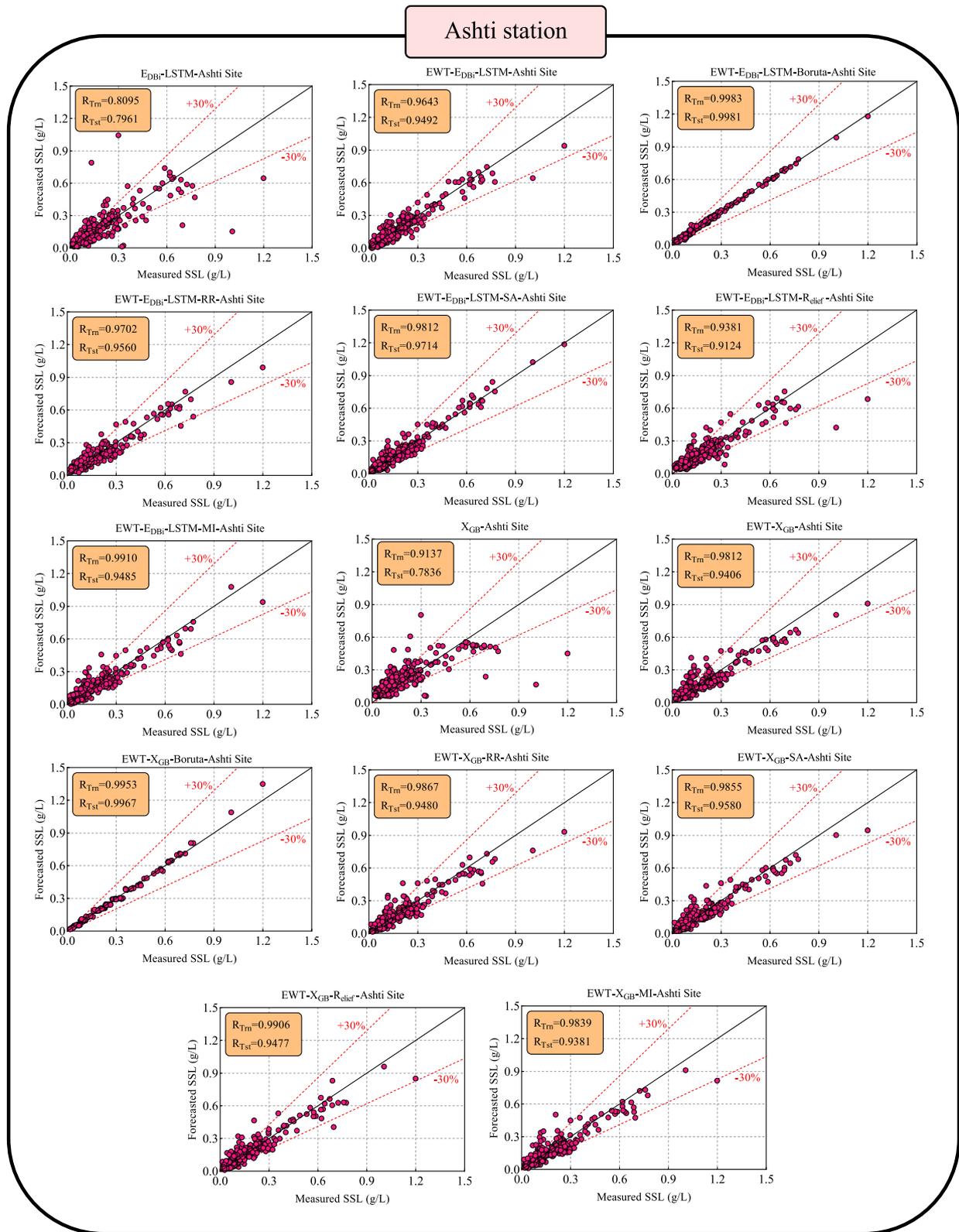
Fig. 13. Performance of all the forecasting models in training and testing phase for Ashti site.

LSTM-R_{elief}, E_{WT}-X_{GB}-R_{elief} models. The good-of-fitness metrics confirmed that the E_{WT}-E_{DBi}-LSTM-Boruta model acquired better precision in relation to all other hybrid versions of the models to forecast daily SSL for the Ashti site. Furthermore, the hybrid version of the models is better in achieving a higher degree of accuracy as compared to the standalone counterpart models. There is an increase in the accuracy

of approximately 20% to 30% in the hybrid models as compared to the standalone models. Overall, the E_{WT}-E_{DBi}-LSTM-Boruta model is the most precise and accurate model to forecast daily SSL for Ashti site.

The hybrid E_{WT}-E_{DBi}-LSTM is better than the E_{WT}-X_{GB} model but could not exceed the E_{WT}-E_{DBI}-LSTM-Boruta model to forecast daily suspended sediment load for the Ashti site (see; Table 5). The standalone
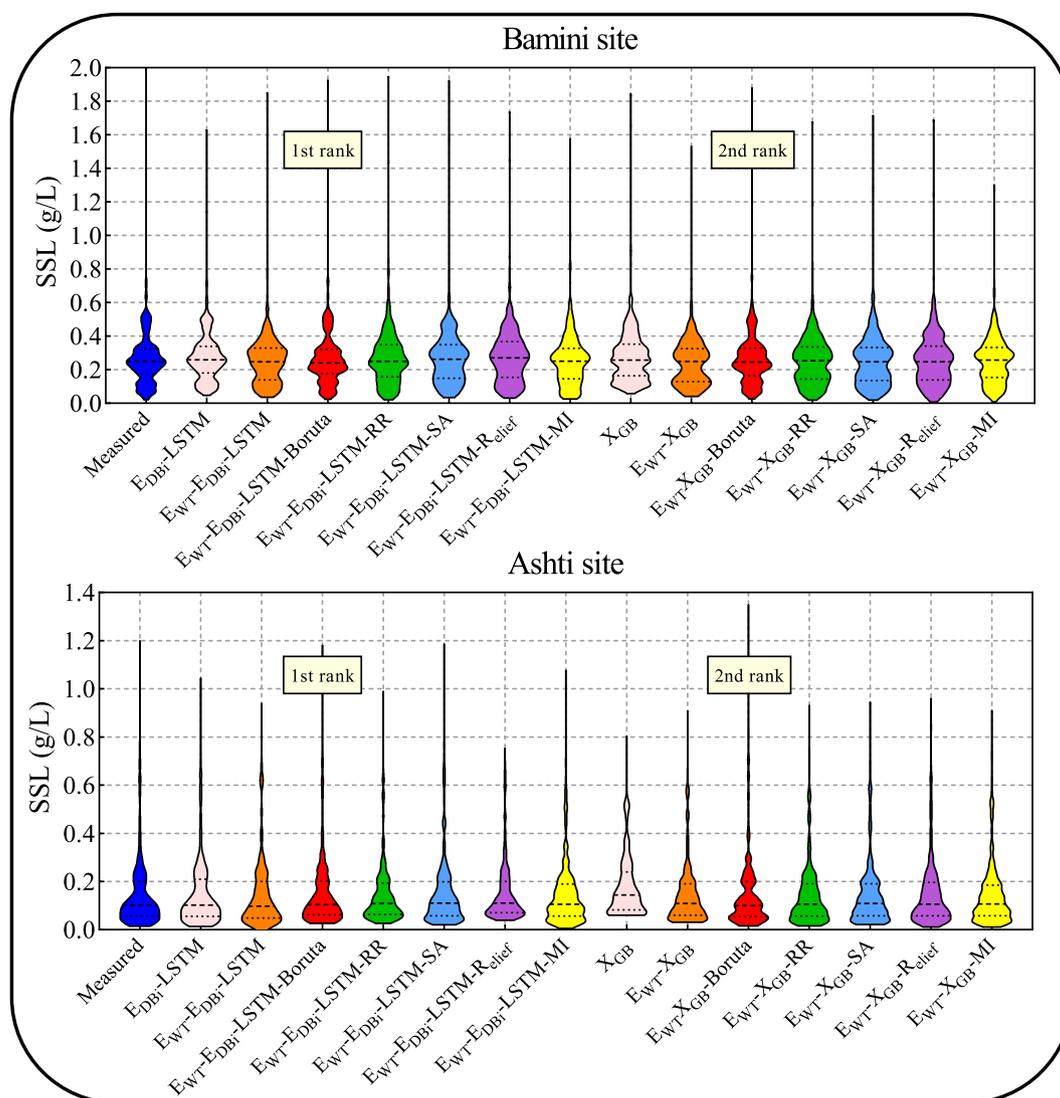
Fig. 14. Comparison of the forecasted *vs.* measured SSL for the testing phase in Bamini and Ashti sites.

E_{DBi}-LSTM and X_{GB} models are again showing poor performance in terms of R, RMSE, RAE, NSE, KGE, and $U_{95\%}$ assessment metrics. By analyzing Table 5, it is cleared that the proposed E_{WT}-E_{DBi}-LSTM-Boruta model for the Ashti site portrays better performance as compared to the other models in this study.

The KGE and $U_{95\%}$ metrics to evaluate the robust performance of E_{WT}-E_{DBi}-LSTM-Boruta *vs.* comparing models for the Ashti site in Fig. 12 shows to forecast daily suspended sediment load. Based on $U_{95\%}$ criteria, the E_{WT}-E_{DBi}-LSTM-Boruta model revealed lower values $U_{95\%} = 0.03$ in comparison with E_{WT}-X_{GB}-Boruta model ($U_{95\%} = 0.04$). On the other hand, a slightly higher KGE (0.97) value was acquired by E_{WT}-X_{GB}-Boruta against E_{WT}-E_{DBi}-LSTM-Boruta (0.96) model in the daily suspended sediment load forecasting. The values of KGE and $U_{95\%}$ for other comparing models can be seen in Fig. 12, confirming that the E_{WT}-E_{DBi}-LSTM-Boruta and E_{WT}-X_{GB}-Boruta models are better to forecast daily suspended sediment load for the Ashti site.

While the scatter plot (Fig. 13) again validated that the accuracy of the E_{WT}-E_{DBi}-LSTM-Boruta model is slightly higher than the E_{WT}-X_{GB}-Boruta model and other comparable models for the Ashti site in the training and testing period based on R for daily forecasting horizon. The E_{WT}-E_{DBi}-LSTM-Boruta ($R_{Trn} = 0.9983$, $R_{Tst} = 0.9981$) model is giving reasonably good preciseness with respect to the benchmark comparing models E_{WT}-X_{GB}-Boruta ($R_{Trn} = 0.9953$, $R_{Tst} = 0.9967$) and so on. Overall, Fig. 13 confirmed that the performance of the E_{WT}-E_{DBi}-LSTM-

Boruta model is better than all other compared models.

Further, assess the comparison of measured *vs.* forecasted SSL generated by the E_{WT}-E_{DBi}-LSTM-Boruta and other benchmarking models in the violin plots in Fig. 14 for the Bamini and Ashti sites. It revealed an obvious distinction between the distributions of forecasted and measured SSL for daily forecasting scenarios. The comparison between the maximum and minimum quartile, first quartile (Q_{25}) and third quartile (Q_{75}), and median (Q_{50}) of forecasted *vs.* measured SSL is given, which established that the E_{WT}-E_{DBi}-LSTM-Boruta model is ranked at 1st place. In contrast, the E_{WT}-X_{GB}-Boruta model is placed at 2nd rank, followed by other comparing models for the Bamini site. Equally, the E_{WT}-E_{DBi}-LSTM-Boruta and E_{WT}-X_{GB}-Boruta models ranked 1st and 2nd places for the Ashti site compared to other daily suspended sediment load (SSL) forecasting.

4.3. Discussion and diagnostic assessment

The Kernel-based residual error distribution and residual error band (REB) of the E_{WT}-E_{DBi}-LSTM-Boruta, E_{WT}-E_{DBi}-LSTM, E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-E_{DBi}-LSTM-R_{elief}, E_{WT}-E_{DBi}-LSTM, E_{DBi}-LSTM, E_{WT}-X_{GB}, E_{WT}-X_{GB}-Boruta, E_{WT}-X_{GB}-MI, E_{WT}-X_{GB}-RR, E_{WT}-X_{GB}-SA, E_{WT}-X_{GB}-R_{elief}, and X_{GB} models are represented in Fig. 15 for the Bamini and Ashti sites in the testing period. The E_{WT}-E_{DBi}-LSTM-Boruta model appeared to be the most accurate model in the
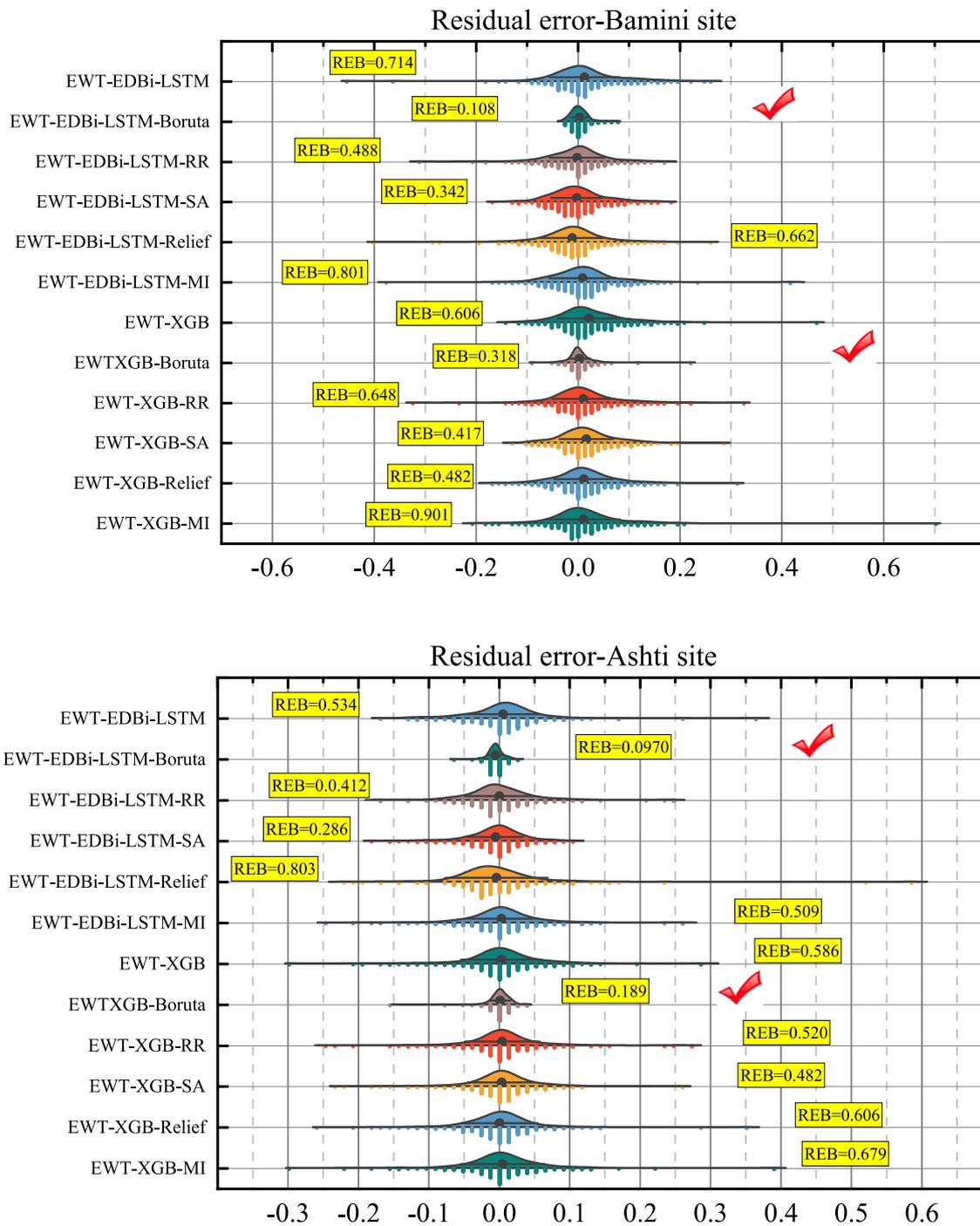
Fig. 15. Kernel-based residual error distribution and residual error band (REB) of all the hybrid models in the testing phase of both sites to daily forecasting of the SSL load in seasonal rivers.

Bamini site by exhibiting the lower value of REB = 0.108, followed by E_{WT}-X_{GB}-Boruta (REB = 0.318), E_{WT}-E_{DBi}-LSTM-SA (REB = 0.342), E_{WT}-X_{GB}-SA (REB = 0.417), E_{WT}-X_{GB}-R_{elief} (REB = 0.482), E_{WT}-E_{DBi}-LSTM-RR (REB = 0.488), E_{WT}-X_{GB} (REB = 0.606), and so on.

Similarly, the E_{WT}-E_{DBi}-LSTM-Boruta model shows better accuracy than other comparing models for the Ashti site (see; Fig. 15) by acquiring the smallest values as E_{WT}-E_{DBI}-LSTM-Boruta (REB = 0.097). The second-best model was E_{WT}-X_{GB}-Boruta (REB = 0.189), followed by E_{WT}-E_{DBi}-LSTM-SA (REB = 0.286), E_{WT}-E_{DBi}-LSTM-RR (REB = 0.412), E_{WT}-X_{GB}-SA (REB = 0.482), E_{WT}-E_{DBi}-LSTM-MI (REB = 0.509) and so on (Fig. 15). In general, the E_{DBI}-LSTM-Boruta model surpassed other benchmarking models for both Bamini and Ashti sites in daily suspended

sediment load forecasting by obtaining the lowest values of REB error distribution.

Fig. 16 describes the time-series plots of forecasted suspended sediment load (i.e., SSL) of the best four models E_{DBI}-LSTM-Boruta, E_{WT}-X_{GB}-Boruta, E_{WT}-E_{DBi}-LSTM-SA, and E_{WT}-X_{GB}-SA against the measured daily SSL for the duration of 17/01/2013 to 10/23/2015 for both sites Bamini and Ashti to draw a direct comparison among the models. The forecast generated by the EDBI-LSTM-Boruta model (red color) appears to be more stable with the measured SSL than EWT-XGB-Boruta E_{WT}-E_{DBi}-LSTM-SA and E_{WT}-X_{GB}-SA models in both Bamini and Ashti sites. Even for the significant fluctuation/spike of measured daily SSL in the periods between 09/13/2015 – 10/03/2015 (Bamini site) and 08/04/2015 –
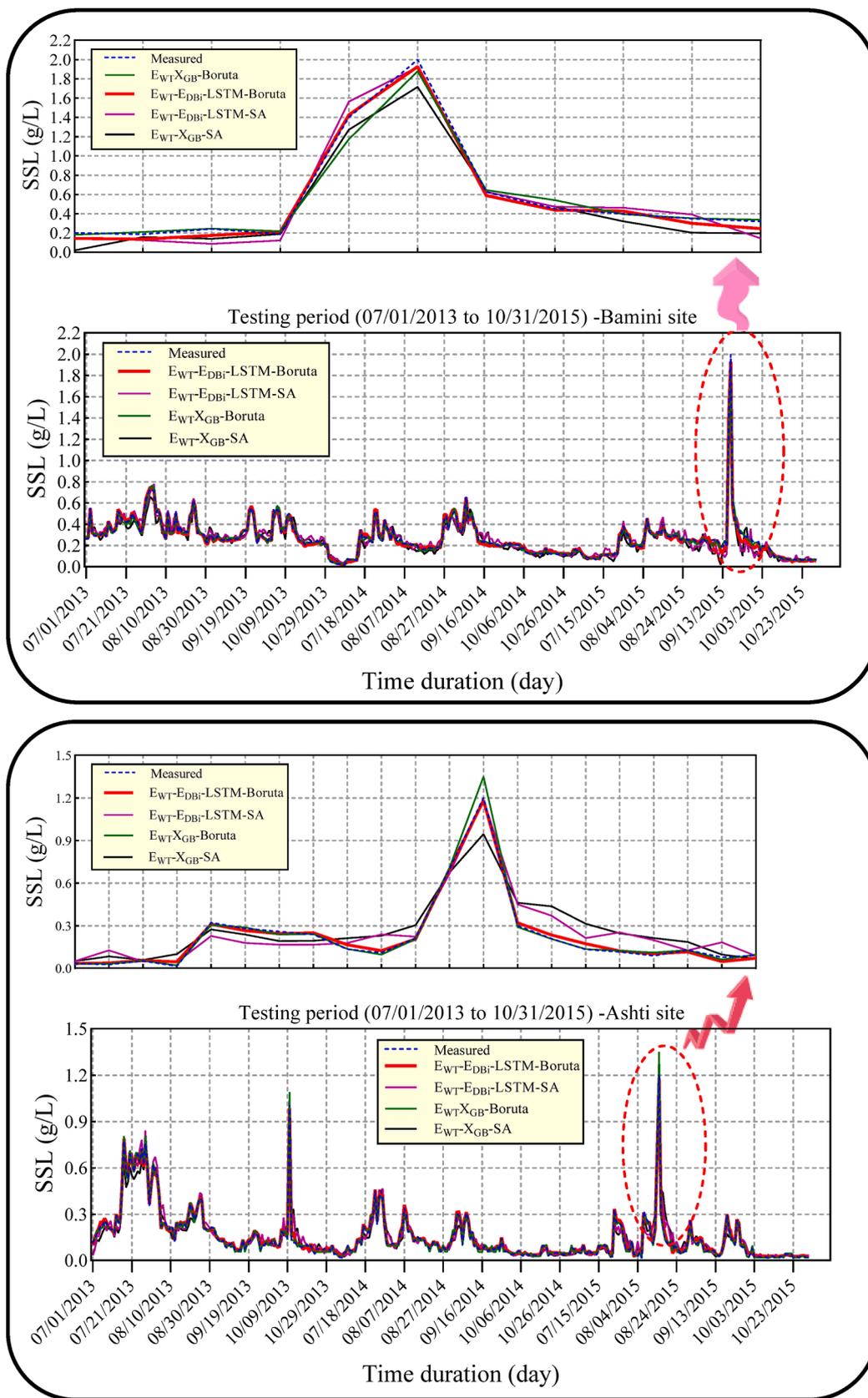
Fig. 16. Trend plot of the superior models to forecast the daily SSL values in Bamini and Ashti sites for the testing period.
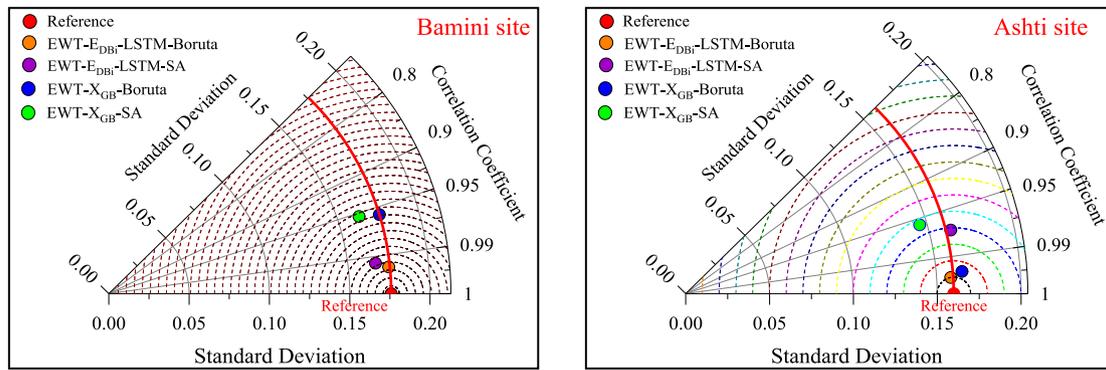
Fig. 17. Taylor diagrams of both sites in the testing phase for the superior models.
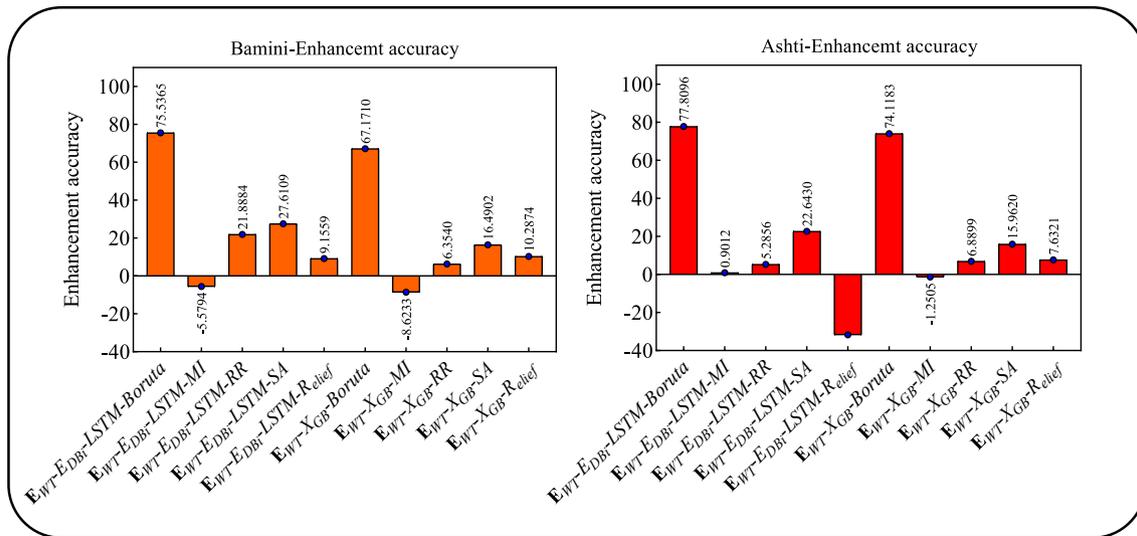
Fig. 18. Enhancement accuracy of all the feature selection-based hybrid models in comparison with the situation containing all decomposed components for both understudy sites.

08/24/2015 (Ashit site), the E_{DBI}-LSTM-Boruta model followed more stability (i.e., accurateness) against the other three benchmarking models. Overall, the E_{DBI}-LSTM-Boruta model achieves better accuracy in terms of time-series plots.

A Taylor diagram is applied to provide A systematic and detailed valuation of the four best models' (E_{DBI}-LSTM-Boruta, E_{WT}-X_{GB}-Boruta, E_{WT}-E_{DBi}-LSTM-SA, and E_{WT}-X_{GB}-SA) competence and performance is described in Taylor diagram (Xu et al., 2016). Fig. 17 demonstrates a more consistent and conclusive relationship between the forecasted and measured/reference SSL using correlation coefficient and standard deviation. The E_{DBI}-LSTM-Boruta model was positioned closely with the reference SSL ranging from 0.95 to 0.99, slightly in proximity with the measured/real SSL E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-X_{GB}-Boruta, and E_{WT}-X_{GB}-SA for the case of Bamini site. Similarly, the E_{DBI}-LSTM-Boruta model appeared to lie more closely with the reference SSL, followed by E_{WT}-X_{GB}-Boruta, E_{WT}-E_{DBi}-LSTM-SA, and E_{WT}-X_{GB}-SA for the Ashti site. This shows that the forecasting accuracy of the E_{DBI}-LSTM-Boruta model was precise at all the sites.

Fig. 18 illustrates the Enhancement accuracy of the feature selection-based hybrid models such as E_{WT}-E_{DBi}-LSTM-Boruta, E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-E_{DBi}-LSTM-R_{elief}, E_{WT}-X_{GB}-Boruta, E_{WT}-X_{GB}-MI, E_{WT}-X_{GB}-RR, E_{WT}-X_{GB}-SA, and E_{WT}-X_{GB}-R_{elief} models for Bamini and Ashti sites. The E_{WT}-E_{DBi}-LSTM-Boruta (75.5365%-Bamini; 77.8096%-Ashti) is the most precise approach based on Enhancement accuracy, followed by E_{WT}-X_{GB}-Boruta (67.1710%-Bamini; 74.1183%-Ashti), E_{WT}-E_{DBi}-LSTM-SA (27.6109%-Bamini;

22.6430%-Ashti) and so on. The Enhancement accuracy in Fig. 18 further ascertains that the feature selection-based E_{WT}-E_{DBi}-LSTM-Boruta model is superior to other feature selection-based models due to the Boruta feature selection technique, which reveals that the best features were selected to acquire optimum accuracy.

5. Conclusion and remarks

The major contribution of this study provides a key and important aspect of determining the best feature selection methods such as Boruta-random forest (Boruta), Relief feature selection (Relief), Mutual information (MI), ridge regression (RR), and Simulated annealing (SA) that can help to improve the forecasting accuracy of the proposed model. Secondly, this research work is aimed by constructing and evaluating a hybridized Empirical Wavelet Decomposition (EWT) technique integrated with the aforementioned feature selection methods (Boruta, Relief, MI, RR, SA) and Encoder decoder Bidirectional-LSTM (E_{DBi}-LSTM), and Extreme learning gradient boosting (X_{GB}) to design the E_{WT}-E_{DBi}-LSTM-Boruta, E_{WT}-E_{DBi}-LSTM-MI, E_{WT}-E_{DBi}-LSTM-RR, E_{WT}-E_{DBi}-LSTM-SA, E_{WT}-E_{DBi}-LSTM-R_{elief}, E_{WT}-X_{GB}-Boruta, E_{WT}-X_{GB}-MI, E_{WT}-X_{GB}-RR, E_{WT}-X_{GB}-SA, and E_{WT}-X_{GB}-R_{elief} algorithms for daily suspended sediment load (SSL) forecasting in Bamini and Ashti sites, India. The goodness-of-fit metrics and visual inspection plots were utilized to verify the accuracy of the proposed AI models for both standalone & hybrid. Assessment of results based on the goodness-of-fit indicators shows that the E_{WT}-E_{DBI}-LSTM-Boruta model outperformed the comparing models for

both the candidate study locations to forecast daily suspended sediment load. Additionally, better outcomes of the hybrid models were found as compared to the standalone version of the models, where a significant reduction in the accuracy of E_{DBi}-LSTM and X_{GB} models was noted during the forecasting. Furthermore, the Boruta feature selection technique was more convincing to determine the best features, which significantly improved the forecasting performance of the E_{WT}-E_{DBI}-LSTM-Boruta model and E_{WT}-X_{GB}-Boruta model in comparison with other benchmarking models. But overall, the E_{WT}-E_{DBI}-LSTM-Boruta model was the top forecasting model in both Bamini and Ashti sites to forecast daily suspended sediment load. Systematically, the outcomes of this work are enormously important and meaningful to design a smart data intelligent model which can accurately perform daily suspended sediment load (SSL) forecasting to monitor water sources, river water quality, irrigation activities, and reservoir operations.

CRediT authorship contribution statement

Mehdi Jamei: Conceptualization, Formal analysis, Visualization, Software, Project administration, Validation. **Mumtaz Ali:** Investigation, Validation. **Anurag Malik:** Formal analysis. **Priya Rai:** Validation, Methodology. **Masoud Karbasi:** Validation, Investigation. **Aitazaz A. Farooque:** Investigation, Supervision. **Zaher Mundher Yaseen:** Investigation, Writing - original draft , Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

References

Abba, S.I., Hadi, S.J., Sammen, S.S., Salih, S.Q., Abdulkadir, R.A., Pham, Q.B., Yaseen, Z. M., 2020. Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. J. Hydrol. 587, 124974 https://doi.org/10.1016/j.jhydrol.2020.124974.

Acikgoz, H., 2022. A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting. Appl. Energy 305, 117912. https://doi.org/10.1016/j.apenergy.2021.117912.

Adams, M.P., Ghisalberti, M., Lowe, R.J., Callaghan, D.P., Baird, M.E., Infantes, E., O'Brien, K.R., 2018. Water residence time controls the feedback between seagrass, sediment and light: Implications for restoration. Adv. Water Resour. 117, 14–26.

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. 541, 902–913.

Ahmed, A.A.M., Deo, R.C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., Yang, L., 2021. Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. J. Hydrol. 599, 126350.

Akay, A.E., Erdas, O., Reis, M., Yuksel, A., 2008. Estimating sediment yield from a forest road network by using a sediment prediction model and GIS techniques. Build. Environ. 43 (5), 687–695.

Ali, M., Prasad, R., Xiang, Y., Khan, M., Ahsan Farooque, A., Zong, T., Yaseen, Z.M., 2021. Variational mode decomposition based random forest model for solar radiation forecasting: New emerging machine learning technology. Energy Reports 7, 6700–6717.

Amjady, N., Keynia, F., 2009. Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. Energy Convers. Manag. 50 (12), 2976–2982.

Armanini, A., Cavedon, V., Righetti, M., 2015. A probabilistic/deterministic approach for the prediction of the sediment transport rate. Adv. Water Resour. 81, 10–18.

Awadh, S., Yousif, A.A., 2021. Lead sorption from aqueous solutions by kaolinite: laboratory experiments. Knowledge-Based Eng. Sci. 2, 1–8.

Azamathulla, H.M., Ab. Ghani, A., Fei, S.Y., 2012. ANFIS-based approach for predicting sediment transport in clean sewer. Applied Soft Computing 12 (3), 1227–1230.

Baniya, M.B., Asaeda, T., K.C., S., Jayashanka, S.M.D.H., 2019. Hydraulic Parameters for Sediment Transport and Prediction of Suspended Sediment for Kali Gandaki River Basin, Himalaya, Nepal. Water 11, 1229. Doi: 10.3390/w11061229.

Betrie, G.D., Mohamed, Y.A., van Griensven, A., Srinivasan, R., 2011. Sediment management modelling in the Blue Nile Basin using SWAT model. Hydrol. Earth Syst. Sci. 15, 807–818.

Bhagat, S.K., Tiyasha, Bekele, D.N., 2018. Economical Approaches for the Treatment and Reutilization of Laundry Wastewater - a Review. J. Ind. Pollut. Control 34, 2164–2178.

Bhagat, S.K., Tung, T.M., Yaseen, Z.M., 2020. Heavy metal contamination prediction using ensemble model: Case study of Bay sedimentation. Australia. J. Hazard. Mater. 403, 123492 https://doi.org/10.1016/j.jhazmat.2020.123492.

Bhagat, S.K., Tiyasha, 2013. Impact of millions of tones of effluent of textile industries: analysis of textile industries effluents in Bhilwara and an approach with bioremediation. Int. J. ChemTech Res. 5, 1289–1298.

Bhattarai, A., Qadir, D., Sunusi, A.M., Getachew, B., Mallah, A.R., 2023. Dynamic Sliding Window-Based Long Short-Term Memory Model Development for Pan Evaporation Forecasting. Knowledge-Based Eng. Sci. 4, 37–54.

Bian, C., He, H., Yang, S., Huang, T., 2020. State-of-charge sequence estimation of lithium-ion battery based on bidirectional long short-term memory encoder-decoder architecture. J. Power Sources 449, 227558. https://doi.org/10.1016/j.jpowsour.2019.227558.

Brédy, J., Gallichand, J., Celicourt, P., Gumiere, S.J., 2020. Water table depth forecasting in cranberry fields using two decision-tree-modeling approaches. Agricult. Water Manage. 233, 106090.

Buyukyildiz, M., Kumcu, S.Y., 2017. An Estimation of the Suspended Sediment Load Using Adaptive Network Based Fuzzy Inference System, Support Vector Machine and Artificial Neural Network Models. Water Resour. Manag. 31, 1343–1359. https://doi.org/10.1007/s11269-017-1581-1.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794.

Chen, Z., Zhu, Z., Jiang, H., Sun, S., 2020. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. J. Hydrol. 591, 125286 https://doi.org/10.1016/j.jhydrol.2020.125286.

Chiang, J.-L., Tsai, Y.-S., 2011. Suspended sediment load estimate using support vector machines in Kaoping river basin, in: Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference On. pp. 1750–1753.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1724–1734. https://doi.org/10.3115/v1/D14-1179.

Cigizoglu, H.K., 2004. Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons. Adv. Water Resour. 27 (2), 185–195.

de Vente, J., Poesen, J., Verstraeten, G., 2005. The application of semi-quantitative methods and reservoir sedimentation rates for the prediction of basin sediment yield in Spain. J. Hydrol. 305 (1-4), 63–86.

Debuse, J.C.W., Rayward-Smith, V.J., 1997. Feature Subset Selection within a Simulated Annealing Data Mining Algorithm. J. Intell. Inf. Syst. 9, 57–81. https://doi.org/10.1023/A:1008641220268.

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2016. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stoch. Environ. Res. Risk Assess. 31, 1211–1240. https://doi.org/10.1007/s00477-016-1265-z.

Elsayed, S., Gupta, M., Chaudhary, G., Taneja, S., Gaur, H., Gad, M., Hamdy Eid, M., Kovács, A., Péter, S., Gaagai, A., Schmidhalter, U., 2023. Interpretation the Influence of Hydrometeorological Variables on Soil Temperature Prediction Using the Potential of Deep Learning Model. Knowledge-Based Eng. Sci. 4, 55–77. https://doi.org/10.51526/kbes.2023.4.1.55-77.

Evaristo, J., McDonnell, J.J., 2019. Global analysis of streamflow response to forest management. Nature. https://doi.org/10.1038/s41586-019-1306-0.

Fadaee, M., Mahdavi-Meymand, A., Zounemat-Kermani, M., 2020. Suspended Sediment Prediction using Integrative Soft Computing Models: On the Analogy between the Butterfly Optimization and Genetic Algorithms. Geocarto Int. 37 (4), 961–977.

Feigl, M., Lebiedzinski, K., Herrnegger, M., Schulz, K., 2021. Machine learning methods for stream water temperature prediction. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/hess-2020-670.

Feng, D., Fang, K., Shen, C., 2020. Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales. Water Resour. Res. 56 (9) https://doi.org/10.1029/2019WR026793.

Ferreira, L.B., da Cunha, F.F., 2020. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. Agricult. Water Manage. 234, 106113.

Gilles, J., 2013. Empirical Wavelet Transform. IEEE Trans. Signal Process. 61, 3999–4010. https://doi.org/10.1109/tsp.2013.2265222.

Goyal, M.K., 2014. Modeling of Sediment Yield Prediction Using M5 Model Tree Algorithm and Wavelet Regression. Water Resour. Manag. 28, 1991–2003. https://doi.org/10.1007/s11269-014-0590-6.

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks 18, 602–610. https://doi.org/10.1016/j.neunet.2005.06.042.

Greig, S.M., Sear, D.A., Carling, P.A., 2005. The impact of fine sediment accumulation on the survival of incubating salmon progeny: Implications for sediment management. Sci. Total Environ. 344 (1-3), 241–258.

Guo, R., Zhao, Z., Wang, T., Liu, G., Zhao, J., Gao, D., 2020. Degradation State Recognition of Piston Pump Based on ICEEMDAN and XGBoost. Appl. Sci. 10, 6593. https://doi.org/10.3390/app10186593.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hassan, M., Shamim, M.A., Sikandar, A., Mehmood, I., Ahmed, I., Ashiq, S.Z., Khitab, A., 2015. Development of sediment load estimation models by using artificial neural networking techniques. Environ. Monit. Assess. 187, 686.

Hassanpour, F., Sharifazari, S., Ahmadaali, K., Mohammadi, S., Sheikhalipour, Z., 2019. Development of the FCM-SVR Hybrid Model for Estimating the Suspended Sediment Load. KSCE J. Civ. Eng. 23, 2514–2523. https://doi.org/10.1007/s12205-019-1693-7.

Hochreiter, S., Schmidhuber, J.J., 1997. Long short-term memory. Neural Comput. 9, 1–32. https://doi.org/10.1162/neco.1997.9.8.1735.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.

Hu, J., Wang, J., 2015. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. Energy 93, 1456–1466. https://doi.org/10.1016/j.energy.2015.10.041.

Hu, J., Wang, J., Ma, K., 2015. A hybrid technique for short-term wind speed prediction. Energy 81, 563–574. https://doi.org/10.1016/j.energy.2014.12.074.

Hu, A., Zhang, K., 2018. Using Bidirectional Long Short-Term Memory Method for the Height of F2 Peak Forecasting from Ionosonde Measurements in the Australian Region. Remote Sens. 10, 1658. https://doi.org/10.3390/rs10101658.

Hur, J.-H., Ihm, S.-Y., Park, Y.-H., 2017. A Variable Impacts Measurement in Random Forest for Mobile Cloud Computing. Wirel. Commun. Mob. Comput. 2017, 1–13. https://doi.org/10.1155/2017/6817627.

Jamei, M., Ahmadianfar, I., Karbasi, M., Jawad, A.H., Farooque, A.A., Yaseen, Z.M., 2021. The assessment of emerging data-intelligence technologies for modeling Mg+2 and SO4− 2 surface water quality. J. Environ. Manage. 300, 113774.

Jamei, M., Ali, M., Malik, A., Karbasi, M., Sharma, E., Yaseen, Z.M., 2022a. Air quality monitoring based on chemical and meteorological drivers: Application of a novel data filtering-based hybridized deep learning model. J. Clean. Prod. 374, 134011.

Jamei, M., Maroufpoor, S., Aminpour, Y., Karbasi, M., Malik, A., Karimi, B., 2022b. Developing hybrid data-intelligent method using Boruta-random forest optimizer for simulation of nitrate distribution pattern. Agric. Water Manag. 270, 107715.

Jamei, M., Karbasi, M., Ali, M., Malik, A., Chu, X., Yaseen, Z.M., 2023a. A novel global solar exposure forecast model based on air temperature: Designing a new multi-processing ensemble deep learning paradigm. Expert Syst. Appl. 119811.

Jamei, Mehdi, Ali, M., Karbasi, M., Sharma, E., Jamei, Mozhdeh, 2023b. Engineering Applications of Artificial Intelligence A high dimensional features-based cascaded forward neural network coupled with MVMD and Boruta-GBDT for multi-step ahead forecasting of surface soil moisture. Eng. Appl. Artif. Intell. 120, 105895. https://doi.org/10.1016/j.engappai.2023.105895.

Jeong, I.-S., Kim, H.-K., Kim, T.-H., Lee, D.H., Kim, K.J., Kang, S.-H., 2016. A Feature Selection Approach Based on Simulated Annealing for Detecting Various Denial of Service Attacks. Softw. Netw. 2016 (1), 173–190.

Khosravi, K., Daggupati, P., Alami, M.T., Awadh, S.M., Ghareb, M.I., Panahi, M., Pham, B.T., Rezaie, F., Qi, C., Yaseen, Z.M., 2019. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. Comput. Electron. Agric. 167, 105041.

Khosravi, K., Mao, L., Kisi, O., Yaseen, Z.M., Shahid, S., 2018. Quantifying hourly suspended sediment load using data mining models: Case study of a glacierized Andean catchment in Chile. J. Hydrol. 567, 165–179. https://doi.org/10.1016/j.jhydrol.2018.10.015.

Kim, H.I., Kim, B.H., 2020. Flood Hazard Rating Prediction for Urban Areas Using Random Forest and LSTM. KSCE J. Civ. Eng. 24 (12), 3884–3896.

Kira, K., Rendell, L.A., 1992. A practical approach to feature selection, in: Machine Learning Proceedings 1992. Elsevier, pp. 249–256.

Kisi, O., Choubin, B., Deo, R.C., Yaseen, Z.M., 2019. Incorporating synoptic-scale climate signals for streamflow modelling over the Mediterranean region using machine learning models. Hydrol. Sci. J. 64 (10), 1240–1252.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13.

Kursa, M.B., Jankowski, A., Rudnicki, W.R., 2010. Boruta – A System for Feature Selection. Fundam. Informaticae 101, 271–285. https://doi.org/10.3233/FI-2010-288.

Kwak, N., Choi, C.-H., 2002. Input feature selection by mutual information based on Parzen window. IEEE Trans. Pattern Anal. Mach. Intell. 24, 1667–1671.

Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., Dech, S., Wegmann, M., 2012. Modelling Forest α-Diversity and Floristic Composition — On the Added Value of LiDAR plus Hyperspectral Remote Sensing. Remote Sens. 4, 2818–2845. https://doi.org/10.3390/rs4092818.

Li, J., Tran, M., Siwabessy, J., 2016. Selecting Optimal Random Forest Predictive Models: A Case Study on Predicting the Spatial Distribution of Seabed Hardness. PLoS One 11, e0149089. Doi: 10.1371/journal.pone.0149089.

Li, T., Li, T.-J., 2018. Sediment transport processes in the Pearl River Estuary as revealed by grain-size end-member modeling and sediment trend analysis. Geo-Marine Lett. 38 (2), 167–178.

Liu, Q., Yang, J., Zhang, K., 2021. An Improved Empirical Wavelet Transform and Sensitive Components Selecting Method for Bearing Fault. Measurement 110348.

Liu, T., Luo, Z., Huang, J., Yan, S., 2018. A comparative study of four kinds of adaptive decomposition algorithms and their applications. Sensors (Switzerland) 18, 1–51. https://doi.org/10.3390/s18072120.

Livieris, I.E., Pintelas, E., Pintelas, P., 2020. A CNN–LSTM model for gold price time-series forecasting. Neural Comput. Appl. 32, 17351–17360. https://doi.org/10.1007/s00521-020-04867-x.

Lyu, B., Zhang, Y., Hu, Y., 2017. Improving PM2.5 Air Quality Model Forecasts in China Using a Bias-Correction Framework. Atmosphere (Basel). 8, 147. Doi: 10.3390/atmos8080147.

Malagó, A., Bouraoui, F., Vigiak, O., Grizzetti, B., Pastori, M., 2017. Modelling water and nutrient fluxes in the Danube River Basin with SWAT. Sci. Total Environ. 603-604, 196–218.

Malik, H., Yadav, A.K., 2021. A novel hybrid approach based on relief algorithm and fuzzy reinforcement learning approach for predicting wind speed. Sustain. Energy Technol. Assessments 43, 100920.

Malik, A., Kumar, A., Singh, R.P., 2019. Application of Heuristic Approaches for Prediction of Hydrological Drought Using Multi-scalar Streamflow Drought Index. Water Resour. Manag. 33 (11), 3985–4006.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash-Sutcliffe Efficiency Index. J. Hydrol. Eng. 11, 597–602. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597).

Meiri, R., Zahavi, J., 2006. Using simulated annealing to optimize the feature selection problem in marketing applications. Eur. J. Oper. Res. 171, 842–858. https://doi.org/10.1016/j.ejor.2004.09.010.

Melesse, A.M., Ahmad, S., McClain, M.E., Wang, X., Lim, Y.H., 2011. Suspended sediment load prediction of river systems: An artificial neural network approach. Agricult. Water Manage. 98 (5), 855–866.

Mengel, D.B., 1993. Fundamentals of Soil Cation Exchange Capacity (CEC). Purdue University Cooperative Extension Service, Agronomy Guide AY-238, West Lafayette.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Nguyen, K.D., Guillou, S., Chauchat, J., Barbry, N., 2009. A two-phase numerical model for suspended-sediment transport in estuaries. Adv. Water Resour. 32, 1187–1196. https://doi.org/10.1016/j.advwatres.2009.04.001.

Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J. Hydrol. 586, 124901.

Nourani, V., Alizadeh, F., Roushangar, K., 2016. Evaluation of a Two-Stage SVM and Spatial Statistics Methods for Modeling Monthly River Suspended Sediment Load. Water Resour. Manag. 30, 393–407. https://doi.org/10.1007/s11269-015-1168-7.

Park, H., Kim, K., Lee, D.K., 2019. Prediction of severe drought area based on random forest: Using satellite image and topography data. Water (Switzerland) 11 (4), 705.

Patino, C.M., Ferreira, J.C., 2015. Confidence intervals: a useful statistical tool to estimate effect sizes in the real world. J. Bras. Pneumol. 41 (6), 565–566.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2019. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. CATENA 177, 149–166. https://doi.org/10.1016/j.catena.2019.02.012.

Rehamnia, I., Benlaoukli, B., Heddam, S., 2020. Modeling of Seepage Flow Through Concrete Face Rockfill and Embankment Dams Using Three Heuristic Artificial Intelligence Approaches: a Comparative Study. Environ. Process. 7 (1), 367–381.

Sadeghi, S.H., Singh, V.P., 2017. Dynamics of suspended sediment concentration, flow discharge and sediment particle size interdependency to identify sediment source. J. Hydrol. 554, 100–110.

Senthil Kumar, A.R., Ojha, C.S.P., Goyal, M.K., Singh, R.D., Swamee, P.K., 2012. Modeling of Suspended Sediment Concentration at Kasol in India Using ANN, Fuzzy Logic, and Decision Tree Algorithms. J. Hydrol. Eng. 17 (3), 394–404.

Shahsavar, A., Jamei, M., Karbasi, M., 2021. Experimental evaluation and development of predictive models for rheological behavior of aqueous Fe3O4 ferrofluid in the presence of an external magnetic field by introducing a novel grid optimization based-Kernel ridge regression supported by sensitivity. Powder Technol. https://doi.org/10.1016/j.powtec.2021.07.037.

Shamaei, E., Kaedi, M., 2016. Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. Appl. Soft Comput. J. 45, 187–196. https://doi.org/10.1016/j.asoc.2016.03.009.

Sharafati, A., Haghbin, M., Tiwari, N.K., Bhagat, S.K., Al-Ansari, N., Chau, K.-W., Yaseen, Z.M., 2021. Performance evaluation of sediment ejector efficiency using hybrid neuro-fuzzy models. Eng. Appl. Comput. Fluid Mech. 15 (1), 627–643.

Shiri, J., Kişi, Ö., 2012. Estimation of Daily Suspended Sediment Load by Using Wavelet Conjunction Models. J. Hydrol. Eng. 17, 986–1000. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000535.

Shojaeezadeh, S.A., Nikoo, M.R., McNamara, J.P., AghaKouchak, A., Sadegh, M., 2018. Stochastic modeling of suspended sediment load in alluvial rivers. Adv. Water Resour. 119, 188–196.

Sikorska-Senoner, A.E., Quilty, J.M., 2021. A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. Environ. Model. Softw. 143, 105094.

Simons, D.B., Şentürk, F., 1992. Sediment transport technology: water and sediment dynamics. Water Resources Publication.

Singh, U.K., Jamei, M., Karbasi, M., Malik, A., Pandey, M., 2022. Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. J. Hydrol. 607, 127549.

Sinha, R., Gupta, A., Mishra, K., Tripathi, S., Nepal, S., Wahid, S.M., Swarnkar, S., 2019. Basin-scale hydrology and sediment dynamics of the Kosi river in the Himalayan foreland. J. Hydrol. 570, 156–166.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9, 1–11. https://doi.org/10.1186/1471-2105-9-307.

Suif, Z., Fleifle, A., Yoshimura, C., Saavedra, O., 2016. Spatio-temporal patterns of soil erosion and suspended sediment dynamics in the Mekong River Basin. Sci. Total Environ. 568, 933–945.

Tao, H., Keshtegar, B., Yaseen, Z.M., 2019. The Feasibility of Integrative Radial Basis M5Tree Predictive Model for River Suspended Sediment Load Simulation. Water Resour. Manag. 33, 4471–4490. https://doi.org/10.1007/s11269-019-02378-6.

Tao, H., Al-Khafaji, Z.S., Qi, C., Zounemat-Kermani, M., Kisi, O., Tiyasha, T., Chau, K.-W., Nourani, V., Melesse, A.M., Elhakeem, M., Farooque, A.A., Pouyan Nejadhashemi, A., Khedher, K.M., Alawi, O.A., Deo, R.C., Shahid, S., Singh, V.P., Yaseen, Z.M., 2021a. Artificial intelligence models for suspended river sediment prediction: state-of the art, modeling framework appraisal, and proposed future research directions. Eng. Appl. Comput. Fluid Mech. 15, 1585–1612. https://doi.org/10.1080/19942060.2021.1984992.

Tao, Y., Liu, C., Liu, C., Zhao, X., Hu, H., 2021b. Empirical wavelet transform method for GNSS coordinate series denoising. J. Geovisualization Spat. Anal. 5, 1–7.

Tao, H., Salih, S., Oudah, A.Y., Abba, S.I., Ameen, A.M.S., Awadh, S.M., Alawi, O.A., Mostafa, R.R., Surendran, U.P., Yaseen, Z.M., 2022. Development of new computational machine learning models for longitudinal dispersion coefficient determination: case study of natural streams. Environ Sci Pollut Res 29 (24), 35841–35861.

Tian, W., Ren, Y., Dong, Y., Wang, S., Bu, L., 2019. Fault monitoring based on mutual information feature engineering modeling in chemical process. Chinese J. Chem. Eng. 27 (10), 2491–2497.

Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res. 49, 6486–6507. https://doi.org/10.1002/wrcr.20517.

Urbanowicz, R.J., Meeker, M., Cava, W.L., Olson, R.S., Moore, J.H., 2018. Relief-based feature selection : Introduction and review. J. Biomed. Inform. 85, 189–203. https://doi.org/10.1016/j.jbi.2018.07.014.

Vergara, J.R., Estévez, P.A., 2014. A review of feature selection methods based on mutual information. Neural Comput. Appl. 24 (1), 175–186.

Verstraeten, G., Poesen, J., de Vente, J., Koninckx, X., 2003. Sediment yield variability in Spain: a quantitative and semiqualitative analysis using reservoir sedimentation rates. Geomorphology 50 (4), 327–348.

Walling, D.E., 1977. Assessing the accuracy of suspended sediment rating curves for a small basin. Water Resour. Res. 13, 531–538. https://doi.org/10.1029/wr013i003p00531.

Walling, D.E., Collins, A.L., 2008. The catchment sediment budget as a management tool. Environ. Sci. Policy 11 (2), 136–143.

Williams, J.R., Berndt, H.D., 1976. Sediment yield prediction based on watershed hydrology. American Society of Agricultural Engineering.

Wu, Y., 2020. Can't Ridge Regression Perform Variable Selection? Technometrics 63, 263–271. https://doi.org/10.1080/00401706.2020.1791254.

Xiang, Z., Yan, J., Demir, I., 2020. A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. Water Resour. Res. 56 https://doi.org/10.1029/2019WR025326.

Xu, Z., Hou, Z., Han, Y., Guo, W., 2016. A diagram for evaluating multiple aspects of model performance in simulating vector fields. Geosci. Model Dev. 9, 4365–4380. https://doi.org/10.5194/gmd-9-4365-2016.

Yaseen, Z.M., Mohtar, W.H.M.W., Ameen, A.M.S., Ebtehaj, I., Razali, S.F.M., Bonakdari, H., Salih, S.Q., Al-Ansari, N., Shahid, S., 2019a. Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region. IEEE Access 7, 74471–74481.

Yaseen, Z.M., Zigale, T.T., Kumar, R., Salih, S.Q., Awasthi, S., Tung, T.M., Al-Ansari, N., Bhagat, S.K., 2019b. Laundry wastewater treatment using a combination of sand filter, bio-char and teff straw media. Sci. Rep. 9, 1–11.

Yin, J., Deng, Z., Ines, A.V.M., Wu, J., Rasu, E., 2020. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). Agric. Water Manag. 242, 106386 https://doi.org/10.1016/j.agwat.2020.106386.

Zhang, B., Zou, G., Qin, D., Lu, Y., Jin, Y., Wang, H., 2021. A novel Encoder-Decoder model based on read-first LSTM for air pollutant prediction. Sci. Total Environ. 765, 144507 https://doi.org/10.1016/j.scitotenv.2020.144507.

Zheng, H., Yuan, J., Chen, L., 2017. Short-Term Load Forecasting Using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation. Energies 10 (8), 1168.

Zounemat-Kermani, M., Meymand, A.M., Ahmadipour, M., 2018. Estimating incipient motion velocity of bed sediments using different data-driven methods. Appl. Soft Comput. 69, 165–176.