Watershed prioritization and decision-making based on weighted sum analysis, feature ranking, and machine learning techniques

Kishanlal Darji[1] · Dhruvesh Patel[1] · Vinay Vakharia[2] · Jaimin Panchal[2] · Amit Kumar Dubey[3] · Praveen Gupta[3] · Raghavendra P. Singh[4]

Abstract

Prediction and validation of Compound factors for prioritization of watersheds are an essential application using machine learning (ML) techniques in water resource engineering. The current paper proposes a methodology to derive 14 morphometric and 3 topo-hydrological parameters using remote sensing (RS) and geographical information systems (GIS). Compound factor (CF) values are calculated using weighted sum analysis (WSA), ReliefF, and the Pearson correlation coefficient, and the important parameters are identified. Two machine learning models, multilayer perceptron (MLP) and support vector machine (SVM), are utilized to predict CF values. Predication accuracy of ML models is evaluated with three parameters, mean absolute error (MAE), Pearson correlation coefficient (PCC), and root mean square error (RMSE). It is observed that the maximum value of PCC equal to 1 is achieved through ReliefF and SVM, whereas minimum MAE and RMSE are observed with ReliefF and SVM when Tenfold cross-validation is applied. Since ReliefF shows better results, CF values are calculated and applied to create the watershed. The proposed methodology is helpful for accurately predicting CF values and advantageous to allocating the proper watershed, which will be useful for decision-making and implementation of conservation techniques for soil and water.

Introduction

Watershed prioritization is crucial for developing watershed management planning and better land management. It is essential to determine the watershed priority in the semi-arid and arid regions as it helps manage water in ungauged rivers. Spatial prioritization and watershed health help to understand watershed conditions, and it also helps derive

better management strategies in the data-scarce region (Alilou et al. 2019).

Remote sensing (RS) and geographical information systems (GIS) are promising tools for extracting crucial parametrical information from watersheds. The RS and GIS technique helps establish interrelationship with parameters and decide the priority of watershed using the WSA technique (Malik et al. 2019). Many researchers have attempted to prioritize watersheds (Thakkar and Dhiman 2007; Patel et al. 2012, 2015; Samal et al. 2015; Kadam et al. 2017; Memon et al. 2020; Kulimushi et al. 2021a; Sengupta et al. 2021) and utilized different analysis techniques for watershed prioritization such as multicriteria decision analysis (MCDA) (Chowdary et al. 2013; Jaiswal et al. 2015; Samal et al. 2015; Vulević and Dragović 2017), Sediment Yield Index (Khan et al. 2001), Weighted Sum Analysis (WSA) (Aher et al. 2014; Memon et al. 2020), and Principal Component Analysis (PCA) (Farhan et al. 2017; Meshram and Sharma 2017). (Patel et al. 2022) has utilized AHP- and TOPSIS-based subwatershed prioritization to facilitate soil erosion susceptibility analysis of

Responsible Editor: Amjad Kallel

✉ Dhruvesh Patel
dhruvesh.patel@sot.pdpu.ac.in

[1] Civil Engineering Department, Pandit Deendayal Energy University, Gandhinagar 382426, India

[2] Department of Mechanical Engineering, Pandit Deendayal Energy University, 382426 Gandhinagar, India

[3] Space Applications Centre, Indian Space Research Organisation, Ahmedabad 380015, Gujarat, India

[4] Indian Institute of Remote Sensing, Indian Space Research Organisation, Dehradun 248001, India

the Ami river basin. Furthermore, (Yadav et al. 2018) has prioritized subwatersheds for potential groundwater zone using topographic sheet and CARTOSAT data to ameliorate the flash flood and droughts, and (Kulimushi et al. 2021b) has utilized compound parameter index (Ci) and Erosion Hazard Rate Index (EHRI) derived from Analytical Hierarchy Process (AHP) for watershed priority assessment. Out of all prioritization, the WSA technique is the most familiar and prominent method for prioritizing the watersheds in the present era; however, the recent development of ML techniques helps to classify, predict, and forecast the laboratory and computer-simulated data for decision-making. Artificial Neural Network (ANN), support vector machine (SVM), and Random Forest are ML models applied in various engineering applications.

ML is a probabilistic model frequently applied for pattern recognition applications. It has been used for applications like fault diagnosis (Kankar et al. 2011), motor current signature analysis (Singh et al. 2014), compressive strength prediction (Sonebi et al. 2016), EEG (Upadhyay et al. 2016), surface roughness prediction (Patel et al. 2020), fault severity analysis (Akhenia et al. 2021), tool wear rate prediction (Shah et al. 2022), and many more applications, regardless of field.

For evaluating any type of ML model and for knowing how well the model predicts without biasedness, cross-validation is used. In k-fold cross-validation, the data is partitioned randomly into the k number of approximately equal sets, from which the model is trained on k-1 sets and tested on the remaining set. The procedure repeated k number of times, and the final result is the average of all the results obtained by repetition. The main benefit of applying this strategy is that each sample is used for training k-1 times and one time for testing, and the average results are obtained. The regression model is used to develop a mathematical function based on the experimental data in which parameters act as features. It is observed that in a feature vector, prediction capability can be improved with identifications of relevant and irrelevant features. To discard irrelevant features, a feature selection strategy can be used to determine the utility of individual features in a feature vector (Vakharia et al. 2015a, b). Other feature ranking techniques like mutual information (Dave et al. 2020), information gain (Naseriparsa et al. 2014), and ReliefF (Kononenko 1994) are also used to improve the accuracy and prediction of classification and regression problems.

With the availability of data, artificial intelligence techniques have been applied in a variety of civil engineering applications. According to literature research, it is found that machine learning strategies are extremely helpful in predicting and validating the experimental results. Furthermore, for decision-making, a particular algorithm may not be suitable; therefore, in such situations, detailed investigations are needed to assess the utility of ML algorithms for particular applications (Vakharia et al. 2022).

In the present paper, RS and GIS techniques are utilized to derive morphometric parameters and topo-hydrological parameters. Initially, the values and ranks to all 14 morphometric and 3 topo-hydrological parameters are assigned, and the CF values are calculated using well-known WSA techniques. After that, two feature ranking techniques, ReliefF and Pearson correlation coefficient, are applied to determine the weightage of morphometric and topological parameters. The weightage obtained from WSA, ReliefF, and PCC will be used to calculate the CF values. Two well-known ML models, SVM and MLP, with Tenfold cross-validations, are applied to identify the best CF values. The accuracy of ML models is examined with mean absolute error (MAE), correlation coefficient (C_r), and root mean square error (RMSE). Based on better prediction capability, decision-making is done for suitable watershed prioritization. As per the literature survey, most authors applied various techniques like parameter averaging, WSA, AHP, MCDM, and TOPSIS. To the authors' best knowledge, decision-making for watershed prioritization using feature ranking and ML techniques is not explored till now. Therefore, to fill the gap, a novel approach for identifying relevant morphometric parameters and consequently the CF values for watersheds prioritization for the decision-making is executed in the present research. The methodology proposed will be helpful in providing an integrated solution to the decision-maker to restore the watersheds against any uncertain critical conditions. Figure 1 shows the watershed prioritization methodology based on feature ranking and ML.

Prediction using machine learning

In recent years, much attention has been focused on machine learning. One of the key causes that are driving its significance is the fact that machine learning provides a unified framework for introducing intelligent decision-making into many research areas. This is one of the primary driving forces behind its popularity in the industry as well as academics. Machine learning models can learn patterns in data whether the learning is supervised or unsupervised. The model is developed based on input data, which will be further used to predict unseen data.

Support vector machine

Support vector machine, more commonly referred to as SVMs, is an intelligent computational model developed for performing classification and regression analysis. In regression analysis, underlying relationships between the dependent variable Y (CF) and the independent variable X (morphological

Fig. 1 Proposed methodology
for watershed prioritization
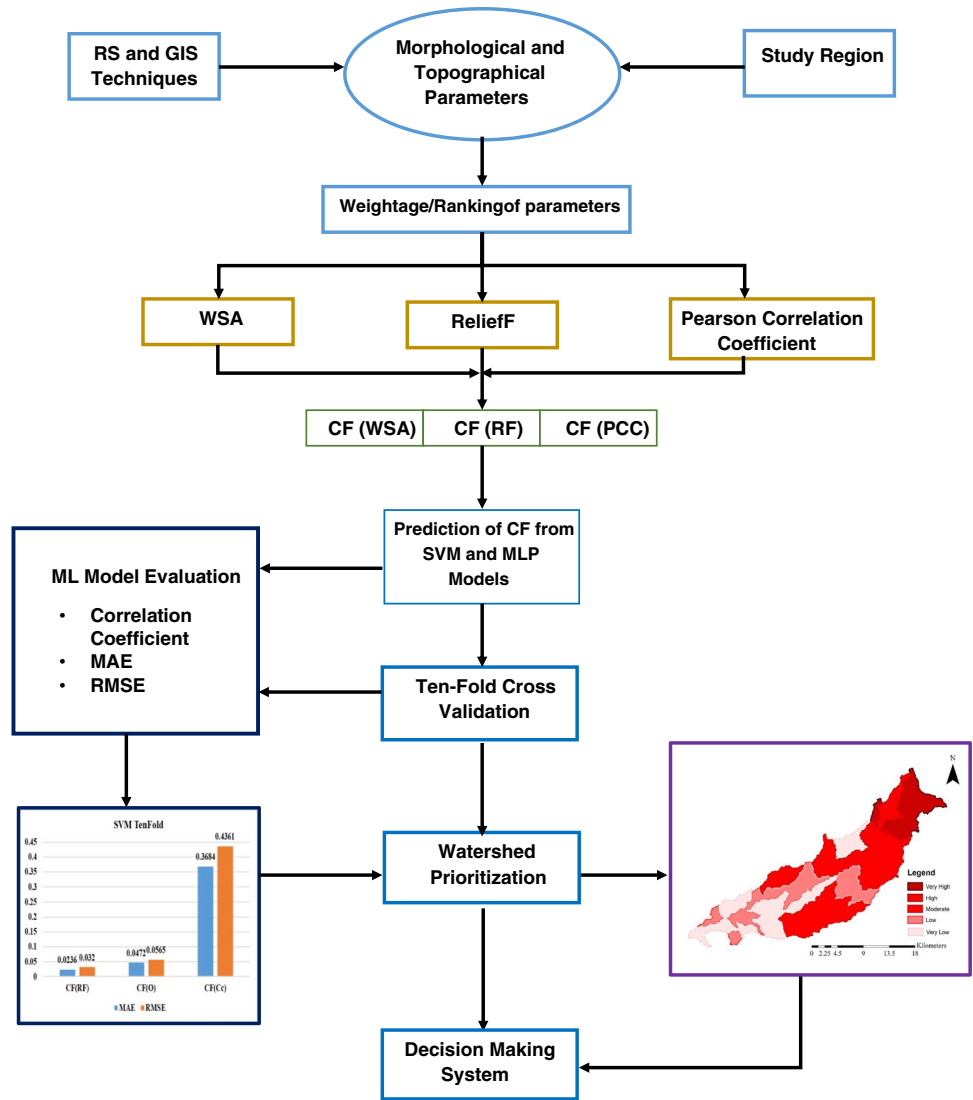
parameters) are determined. The relationship is generally mapped by using a kernel function, which described how the estimated Y deviates from the calculated values. Let the training data be represented as

$$\tau = \left\{ (X_1, Y_1), (X_2, Y_2), \ldots \ldots, (X_n, Y_N) \right\} \in \chi \times \Re \quad (1)$$

where χ represent the space for input patterns. If the kernel function is assumed to be linear, it can be expressed as

$$h(x) = \langle W, X \rangle + b \quad (2)$$

SVM is formulated as a convex optimization problem as

$$\min \frac{1}{2} \|W\|^2 \quad (3)$$

$$\text{s.t.} Y_i - \left(\langle W, X_i \rangle + b \right) \leq \varepsilon, \forall i \quad (4)$$

$$\left(\langle W, X_i + b \rangle - Yi \leq \varepsilon, \forall i \right) \quad (5)$$

Here, w represents the hyperplane orientation, and b is a scalar quantity known as intercept. When margin is considered Eqs. 3, 4, and 5 can be further modified as (Vapnik 1995)

$$\min \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{N} \left(\xi_i + \xi_i^* \right) \quad (6)$$

$$\text{s.t.} Y_i - \left(\langle W, X_i \rangle + b \right) \leq \varepsilon + \xi_i, \forall i \quad (7)$$

$$\left(\langle W, X_i \rangle + b \right) - Y_i \leq \varepsilon + \xi_i^*, \forall i \\ \xi_i, \xi_i^* \geq 0 \quad (8)$$

ξ_i, ξ_i^* represents the slack variables which are introduced to deal with infeasible constraints. In present study, the

penalty parameter "C" is chosen as 10 and radial basis kernel function is used to effectively map the nonlinearity in morphological parameters (Fig. 2a).

Multilayer perceptron (MLP)

Multilayer perceptron (MLP) is another type of ANN architecture which consists of several layers and functions as a global approximator for nonlinear input–output mapping. MLPs are composed of neurons called perceptron. In our model, morphological features which is represented as $\{g = g_1, g_2, g_3, \cdots, g_n\}$ become an input to perceptron. The input morphological features are passed through a function u which computes the weighted sum of input features as follows:

The output of the perceptron is calculated by an activation function f, which is represented as

$$Y = f(u(X)) = \begin{cases} 1, if u(X) > \theta \\ 0, otherwise, \end{cases} \quad (9)$$

The equation $w_1 g_1 + w_2 g_2 + \cdots + w_n g_n - \theta = 0$ represents the equation of hyperplane. Output $= 1$ reflects that input morphological parameters lies above the hyperplane, and Output $= 0$ reflects that input morphological parameters lie below the hyperplane. Hence for output, it gives a set of continuous values and mean square error as a loss function. (Fig. 2b) shows the architecture of MLP with one hidden layer. In the present study, authors have used only one hidden layer based on prior experience.

Fig. 2 Prediction using **a** support vector machine and **b** multilayer perceptron
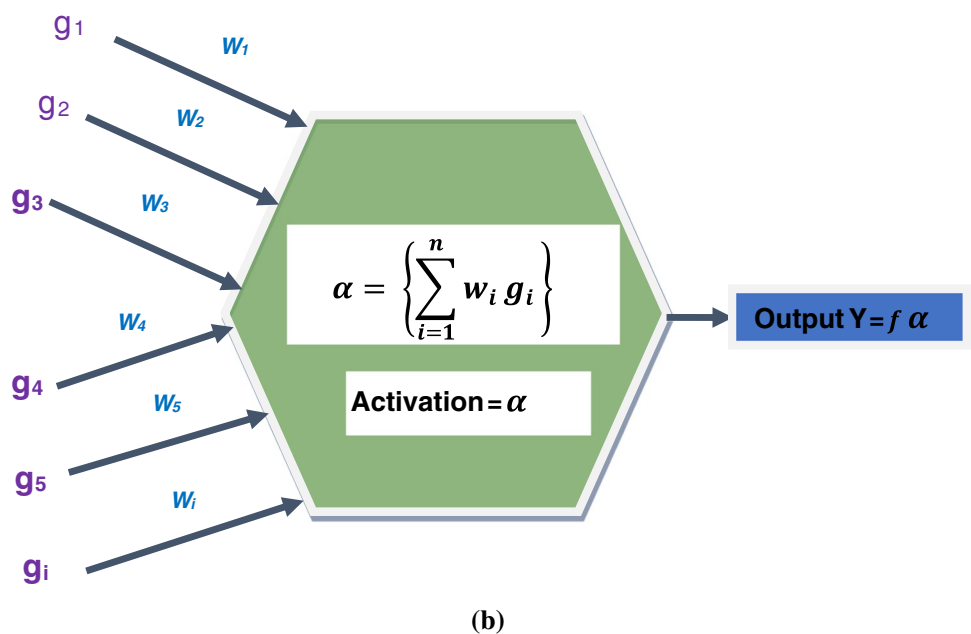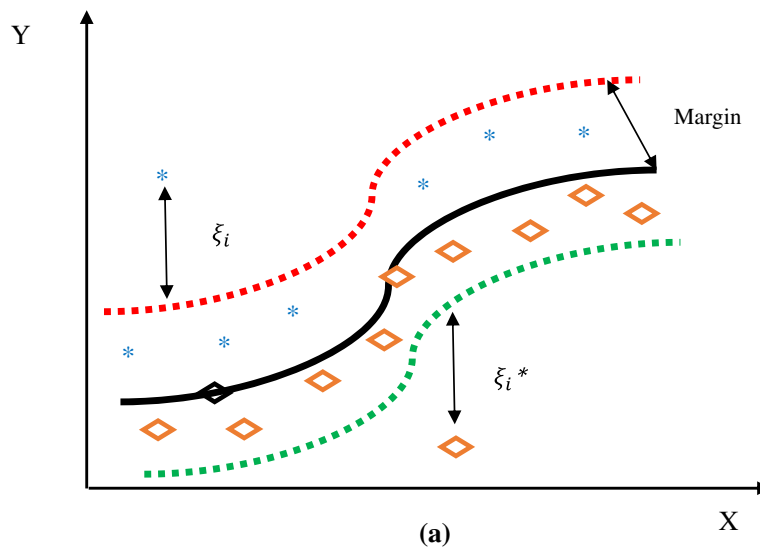
Parameter estimations through GIS and remote sensing

To determine the CF for any catchment area of the River, a Rel-River catchment situated in Banaskantha district of Gujarat, India, is considered in the present study as shown in (Fig. 3a). Basin lies between 24° 50′ N to 24° 75′ N latitude and 72° 00′ E to 72° 45′ E longitude and has an area of 431 km^2 (Memon et al. 2020; Darji et al. 2021). The basin area is partitioned into 51 microwatersheds, and Cartosat DEM 10 m resolution is used to calculate slope and watershed boundary. ArcHydro toolset is used for calculating and deriving the boundary of the watershed. Drainage boundaries are derived using Survey of India (SOI) toposheets (42D02, 42D03, 42D06), and the ordering of the drainages is given according to the proposed technique by (Strahler 1964) (Fig. 3b). ArcGIS 10.5 software was used to generate the boundary of watersheds, digitize drainages, and derive different morphometric parameters. Morphometric parameters such as linear aspects, areal aspects, and relief aspects (Horton 1932, 1945; Miller 1953; Schumm
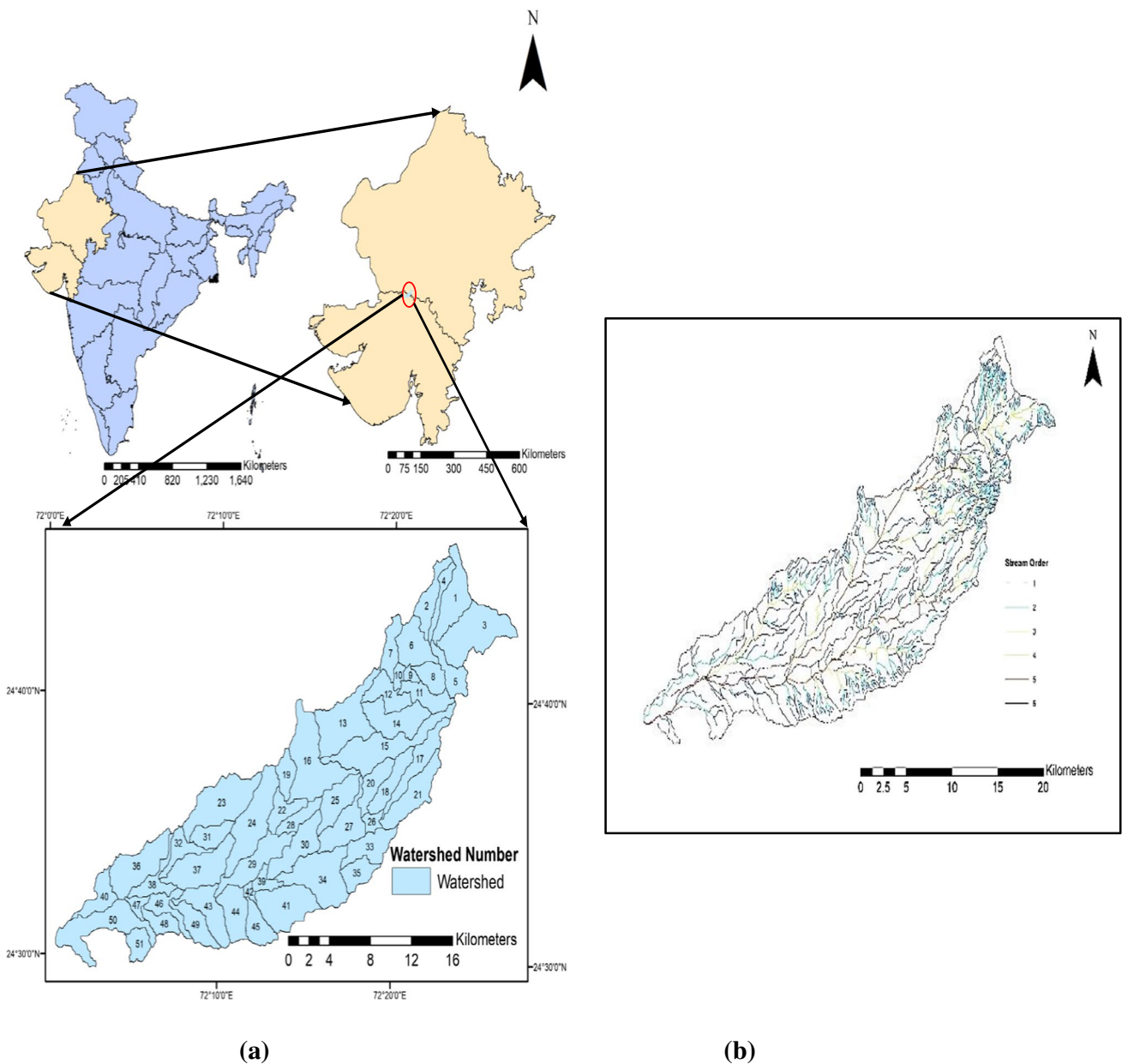

(a)

(b)

Fig. 3 Study area. **a** Rel river basin map. **b** Rel river drainage map

1956; Melton 1957; Faniran 1968; Strahler 1997; Nooka Ratnam et al. 2005) and topo-hydrological parameters such as Stream Power Index (SPI) (Whipple and Tucker 1999), Sediment Transport Index (STI) (Moore and Burch 1986), and Topographic Wetness Index (TWI) (Beven and Kirkby 1979) are considered for the prioritization of watersheds (Table 1). The ranking of the watersheds is calculated based on the WSA technique established by (Aher et al. 2014). Every microwatershed has different and unique characteristics, and watershed priority is carried out to identify the vulnerable areas for erosion. The weights have been calculated based on the correlation between parameters and their values. Its relation does a ranking of the parameters to soil erosion. Based on the literature, the linear factor showed a positive correlation with soil erosion, so maximum priority is given to the higher value, i.e., rank 1. In contrast, shape factors negatively correlated with soil erosion, and the ranking for these parameters was given in reverse order.

The final ranking and prioritization of watersheds have been done using CF values. The CF values are calculated with the help of morphological parameters weightage obtained from WSA, ReliefF, and Pearson correlation coefficient methods.

Table 1 Morphometric and topo-hydrological parameters considered in study

Parameter	Aspect	Parameters	Formula	References
Morphometric parameters	Linear aspect	Stream order (N_u)	Hierarchical order	(Strahler 1964)
		Stream length (L)	Length of stream	(Horton 1945)
		Length of overland flow (L_o)	$L_o = 1/D_d * 2$, where L_o = length of overland flow and D_d = drainage density	(Horton 1945)
		Bifurcation ratio (R_b)	$R_b = N_u/N_{u+1}$, where N_u = total no. of stream segments of order N_{u+1} = number of segments of next higher order	(Schumm 1956)
	Areal aspect	Stream frequency (F_u)	$F_u = N_u/A$, where A = area of the basin	(Horton 1945)
		Drainage density (D_d)	$D_d = L_u/A$, where L_u = total length of all order streams	(Horton 1932)
		Form factor (R_f)	$R_f = A/L_b^2$, where L_b = basin length	(Horton 1932)
		Elongation ratio (R_e)	$R_e = \sqrt{4 \times \frac{A}{p_i}}/L_b$	(Schumm 1956)
		Shape factor (B_s)	$B_s = L_b^2/A$	(Nooka Ratnam et al. 2005)
		Compactness constant (C_c)	$C_c = 0.2821P/A^{0.5}$	(Horton 1945)
		Circularity ratio (R_c)	$R_c = 4\pi A/P^2$, where P = basin perimeter	(Miller 1953)
		Drainage texture (R_t)	$R_t = N_u/P$	(Horton 1945)
		Infiltration number (I_f)	$I_f = F_u \times D_d$	(Faniran 1968)
		Constant of channel maintenance (C)	$C = 1/D_d$	(Schumm 1956)
	Relief aspect	Basin relief (H)	The maximum vertical distance between the highest and lowest points in the watershed	(Schumm 1956)
		Relief ratio (R_h)	$R_h = H/L_b$	(Schumm 1956)
		Ruggedness no. (R_n)	$R_n = H \times D_d$	(Melton 1957)
Topo-hydrological parameters		Topographic Wetness Index (TWI)	$TWI = \ln\left(\frac{a}{\tan(\beta)}\right)$, where a = upslope catchment area divided by contour length and β = slope of the catchment	(Beven and Kirkby 1979)
		Stream Power Index (SPI)	$SPI = (a \times \tan\beta)$	(Whipple and Tucker 1999)
		Sediment Transport Index (STI)	$STI = \left(\frac{a}{22.13}\right)^{0.06} \times \left(\frac{\sin\beta}{0.0896}\right)^{1.3}$	(Moore and Burch 1986)

Table 2 Stream order of different watersheds

WS no	Stream orders						Summation of stream	L (length)	R_b (bifurcation ratio)	Basin relief
	I	II	III	IV	V	VI				
1	36	21	12	5	1	0	75	45.5	2.72	368
2	8	6	0	0	0	0	14	18.6	1.33	202
3	59	31	15	9	0	0	114	52.5	1.88	541
4	14	12	3	0	2	0	31	22.2	2.22	203
5	31	14	13	0	0	0	58	22.5	1.65	493
6	16	6	0	0	10	0	32	18.1	1.63	143
7	23	9	12	0	0	0	44	20.3	1.65	154
8	24	12	9	1	2	0	48	17.2	3.21	162
9	4	1	1	0	4	0	10	4.7	1.75	47
10	5	0	1	0	4	0	10	3.6	2.63	41
11	38	21	15	0	0	0	74	30.8	1.60	170
12	19	8	8	1	6	0	42	22.6	2.89	59
13	5	0	2	0	4	0	11	15.7	1.50	62
14	40	18	4	11	0	0	73	38.8	2.36	80
15	39	16	20	0	0	0	75	44.5	1.62	206
16	20	13	3	0	6	0	42	47.3	2.12	78
17	27	9	6	6	1	0	49	31.0	2.88	135
18	7	4	3	0	0	0	14	14.6	1.54	67
19	7	4	2	0	2	0	15	23.3	1.58	40
20	4	0	5	0	0	0	9	9.2	0.80	59
21	19	12	2	3	5	0	41	34.1	2.21	83
22	1	0	0	0	1	0	2	15.5	1.00	32
23	39	19	11	8	0	0	77	47.5	1.72	73
24	11	4	4	0	1	0	20	36.4	2.58	67
25	8	7	0	0	0	0	15	12.3	1.14	65
26	1	0	1	0	3	0	5	4.2	0.67	23
27	7	3	0	0	4	0	14	11.7	1.54	65
28	2	0	0	0	0	0	2	5.3	2.00	39
29	1	0	1	0	1	1	4	20.2	1.00	44
30	10	4	1	1	9	0	25	22.9	1.90	48
31	3	1	0	1	0	0	5	8.2	2.00	42
32	1	1	0	3	0	0	5	8.6	0.67	38
33	3	3	2	0	0	0	8	12.4	1.25	43
34	28	18	5	6	1	0	58	33.8	3.00	51
35	15	6	7	0	0	0	28	16.4	1.68	30
36	6	3	0	1	0	2	12	22.1	1.83	34
37	5	4	0	1	0	0	10	17.2	2.63	50
38	2	1	0	3	0	0	6	10.8	1.17	30
39	1	0	0	1	2	0	4	7.3	0.75	25
40	1	2	0	0	0	1	4	13.7	1.25	17
41	38	22	10	5	1	0	76	33.9	2.73	50
42	0	0	1	0	2	0	3	4.0	0.50	20
43	6	2	0	0	0	3	11	13.9	1.83	40
44	10	6	0	0	4	1	21	30.5	2.39	27
45	13	7	5	0	1	0	26	15.6	2.75	34
46	3	1	0	0	0	3	7	14.4	1.67	34
47	1	0	0	0	0	2	3	5.4	0.50	10
48	3	0	0	0	0	0	3	4.5	3.00	26
49	1	0	0	0	0	0	1	4.4	1.00	38
50	2	3	0	0	0	3	8	26.8	0.83	58
51	2	1	0	0	0	0	3	5.9	2.00	35

Table 3 Parameter values

WS no	R_b	D_d	F_u	R_t	L_o	R_f	B_s	R_e	C_c	I_f	R_c	C	R_h	R_n	STI	SPI	TWI
1	2.72	4.05	6.67	2.95	0.12	0.42	2.39	0.73	2.14	27.01	0.22	0.25	9.61	1.49	4.55	−0.03	7.81
2	1.33	2.74	2.06	0.62	0.18	0.45	2.23	0.76	2.43	5.66	0.17	0.36	12.98	0.55	2.02	−0.47	8.08
3	1.88	2.84	6.17	3.96	0.18	0.39	2.56	0.71	1.89	17.53	0.28	0.35	10.30	1.54	7.81	0.53	7.54
4	2.22	6.24	8.71	1.59	0.08	0.49	2.05	0.79	2.92	54.38	0.12	0.16	9.14	1.27	2.41	−0.25	7.70
5	1.65	3.95	10.19	3.14	0.13	0.46	2.18	0.76	2.18	40.27	0.21	0.25	21.93	1.95	9.42	1.44	6.23
6	1.63	2.03	3.59	1.75	0.25	0.43	2.32	0.74	1.73	7.28	0.34	0.49	7.90	0.29	1.33	−0.83	9.24
7	1.65	3.78	8.18	2.56	0.13	0.46	2.16	0.77	2.09	30.91	0.23	0.26	7.58	0.58	2.01	−0.32	7.78
8	3.21	3.51	9.80	3.82	0.14	0.47	2.14	0.77	1.60	34.36	0.39	0.29	9.43	0.57	2.30	0.01	7.18
9	1.75	3.65	7.74	1.40	0.14	0.56	1.78	0.85	1.77	28.22	0.32	0.27	9.97	0.17	1.52	−0.90	9.10
10	2.63	2.24	6.28	1.30	0.22	0.55	1.83	0.83	1.73	14.03	0.34	0.45	11.52	0.09	1.73	−0.50	8.81
11	1.60	4.90	11.76	3.63	0.10	0.45	2.21	0.76	2.29	57.55	0.19	0.20	5.52	0.83	2.43	0.26	6.80
12	2.89	3.18	5.92	1.98	0.16	0.45	2.25	0.75	2.25	18.85	0.20	0.31	2.61	0.19	1.39	−0.85	8.74
13	1.50	0.92	0.64	0.33	0.55	0.39	2.53	0.71	2.25	0.59	0.20	1.09	3.94	0.06	0.96	−1.48	9.82
14	2.36	3.36	6.32	2.85	0.15	0.42	2.40	0.73	2.12	21.25	0.22	0.30	2.06	0.27	0.93	−1.15	9.65
15	1.62	2.55	4.30	1.67	0.20	0.39	2.54	0.71	3.03	10.95	0.11	0.39	4.63	0.53	1.06	−1.24	9.37
16	2.12	2.46	2.18	1.25	0.20	0.39	2.57	0.70	2.15	5.36	0.22	0.41	1.65	0.19	0.84	−1.55	10.06
17	2.88	3.59	5.68	1.64	0.14	0.43	2.31	0.74	2.86	20.37	0.12	0.28	4.36	0.48	1.05	−1.23	9.18
18	1.54	2.14	2.06	0.72	0.23	0.45	2.23	0.75	2.12	4.40	0.22	0.47	4.59	0.14	0.65	−1.66	10.27
19	1.58	5.30	3.41	1.20	0.09	0.47	2.11	0.78	1.67	18.06	0.36	0.19	1.72	0.21	0.80	−1.46	9.64
20	0.80	1.55	1.52	0.53	0.32	0.46	2.19	0.76	1.98	2.36	0.25	0.64	6.43	0.09	0.69	−1.63	10.08
21	2.21	3.15	3.79	1.61	0.16	0.42	2.38	0.73	2.18	11.91	0.21	0.32	2.44	0.26	0.68	−1.72	10.10
22	1.00	2.93	0.38	0.12	0.17	0.46	2.16	0.77	2.02	1.11	0.25	0.34	2.07	0.09	1.07	−1.52	9.93
23	1.72	2.71	4.40	2.66	0.18	0.39	2.54	0.71	1.95	11.94	0.26	0.37	1.54	0.20	0.74	−1.68	10.24
24	2.58	2.17	1.19	0.68	0.23	0.40	2.53	0.71	2.04	2.59	0.24	0.46	1.84	0.15	0.66	−1.65	10.32
25	1.14	1.43	1.73	0.78	0.35	0.43	2.31	0.74	1.85	2.47	0.29	0.70	5.27	0.09	0.61	−1.79	10.13
26	0.67	1.98	2.35	0.53	0.25	0.52	1.91	0.82	1.82	4.64	0.30	0.51	5.45	0.05	0.75	−1.66	10.06
27	1.54	1.19	1.42	0.58	0.42	0.43	2.35	0.74	2.17	1.70	0.21	0.84	5.53	0.08	0.85	−1.42	10.13
28	2.00	1.20	0.45	0.12	0.42	0.47	2.11	0.78	2.22	0.54	0.20	0.84	7.34	0.05	0.54	−1.75	10.30
29	1.00	1.83	0.36	0.13	0.27	0.42	2.39	0.73	2.57	0.66	0.15	0.55	2.18	0.08	1.36	−1.56	10.41
30	1.90	2.27	2.47	1.03	0.22	0.42	2.36	0.73	2.16	5.61	0.21	0.44	2.09	0.11	0.81	−1.41	10.10
31	2.00	1.57	0.95	0.35	0.32	0.46	2.16	0.77	1.76	1.49	0.32	0.64	5.10	0.07	0.56	−1.66	10.72
32	0.67	1.69	0.98	0.24	0.30	0.47	2.15	0.77	2.64	1.66	0.14	0.59	4.40	0.06	0.91	−1.78	10.64
33	1.25	1.53	0.99	0.38	0.33	0.44	2.29	0.75	2.08	1.51	0.23	0.65	3.46	0.07	0.66	−1.51	10.49
34	3.00	1.95	3.35	1.86	0.26	0.39	2.54	0.71	2.12	6.55	0.22	0.51	1.51	0.10	0.77	−1.61	10.11
35	1.68	3.08	5.25	1.99	0.16	0.46	2.16	0.77	1.71	16.17	0.34	0.32	1.82	0.09	0.55	−1.71	10.44
36	1.83	1.61	0.87	0.49	0.31	0.41	2.46	0.72	1.86	1.40	0.29	0.62	1.54	0.05	0.67	−1.75	10.69
37	2.63	1.25	0.73	0.40	0.40	0.41	2.46	0.72	1.92	0.91	0.27	0.80	2.91	0.06	0.52	−1.74	10.81
38	1.17	2.03	1.13	0.24	0.25	0.46	2.16	0.77	3.09	2.30	0.11	0.49	2.79	0.06	0.69	−1.94	10.84
39	0.75	2.57	1.42	0.35	0.19	0.50	1.98	0.80	1.93	3.64	0.27	0.39	3.45	0.06	0.93	−1.43	10.11
40	1.25	3.79	1.11	0.33	0.13	0.49	2.05	0.79	1.81	4.19	0.31	0.26	1.24	0.06	0.92	−1.94	10.76
41	2.73	2.40	5.39	3.34	0.21	0.41	2.47	0.72	1.71	12.94	0.34	0.42	1.48	0.12	0.68	−1.78	10.11
42	0.50	3.66	2.73	0.48	0.14	0.57	1.74	0.85	1.70	10.00	0.35	0.27	4.97	0.07	1.60	−1.36	10.50
43	1.83	1.42	1.12	0.46	0.35	0.43	2.35	0.74	2.17	1.59	0.21	0.70	2.87	0.06	0.84	−1.86	10.70
44	2.39	3.59	2.48	1.31	0.14	0.43	2.30	0.74	1.55	8.89	0.41	0.28	0.89	0.10	0.77	−1.78	10.49
45	2.75	2.64	4.41	1.76	0.19	0.46	2.19	0.76	1.72	11.62	0.34	0.38	2.19	0.09	0.56	−1.92	10.43
46	1.67	2.77	1.35	0.31	0.18	0.46	2.15	0.77	2.80	3.73	0.13	0.36	2.36	0.09	0.85	−1.92	10.83
47	0.50	2.62	1.45	0.30	0.19	0.53	1.90	0.82	1.94	3.81	0.26	0.38	1.85	0.03	1.11	−2.20	10.74
48	3.00	1.06	0.71	0.22	0.47	0.48	2.09	0.78	1.85	0.76	0.29	0.94	5.80	0.03	0.38	−1.88	10.74
49	1.00	0.68	0.15	0.05	0.74	0.45	2.22	0.76	2.10	0.10	0.23	1.48	8.58	0.03	0.40	−1.79	11.02
50	0.83	2.36	0.70	0.20	0.21	0.42	2.40	0.73	3.31	1.66	0.09	0.42	2.16	0.14	0.39	−1.76	10.95
51	2.00	1.38	0.70	0.25	0.36	0.48	2.10	0.78	1.63	0.96	0.38	0.73	5.91	0.05	0.89	−0.95	9.99

WSA techniques

The WSA is a statistical approach applied to calculate CF values. The CF values are estimated using cross-correction analysis by morphometric parameters and the assigned weightage. The statistical expression for CF is written as follows (Aher et al. 2014):

$$CF(WSA) = RMP \times WMP \tag{10}$$

where CF (WSA) represents compound factor, RMP represents the rank (preliminary priority) estimated from morphometric and topo-hydrological parameters, and WMP represents the weight of morphometric and topo-hydrological parameters obtained using a cross-correlation study (Table 2, 3, and 4). Weights are estimated with a ratio of morphometric and topo-hydrological parameters with a sum of the correlation coefficient value of each parameter (Table 5). Based on the weightage of parameters, a model is constructed for sorting watershed prioritization, which is computed as follows:

$$\begin{aligned}
CF(WSA) = &(0.055447) \times R_b + (0.087678) \times D_d + (0.114486) \\
&\times F_u + (0.111822) \times R_t + (-0.08424) \times L_o \\
&+ (0.025944) \times R_f + (-0.01748) \times B_s + (0.028237) \\
&\times R_e + (-0.00423) \times C_c + (0.112984) \times I_f \\
&+ (0.007042) \times R_c + (0.089087) \times C + (0.047331) \\
&\times R_h + (0.115931) \times R_n + (0.090761) \times STI \\
&+ (0.105501) \times SPI + (0.113691) \times TWI
\end{aligned} \tag{11}$$

As observed in Table 7 and Eq. 11, the highest CF value is 51.13, and it is for watershed 49, the second-highest value is 47.48, and it is for watershed 48, and so on for other watersheds, as shown in (Fig. 4a).

ReliefF

ReliefF is a feature ranking method for selecting important features for classification and regression problems. In this method, predictors that give different values for the same response values are penalized, and predictors that give different values for different response values are rewarded. Final predictor weights are computed using intermediate weights.

$$P_w = \frac{W_{Rp}}{W_R} - \frac{p_R - W_{Rp}}{n - W_R} \tag{12}$$

where n is several instances, W_{RP} is weighted with different response values and different values for predictor, W_R is weighted with different values for response R, and P_R is weights with different values for predictor p (Robnik-Šikonja et al. 1997).

The ReliefF method is used to assign the ranks to the morphometric parameters. I_f ranked first with the weightage of 0.058, whereas F_u is ranked second with the weightage of 0.056. Other subsequent values, ranks, and weights are mentioned in Table 6. The final equation of compound factor obtained using the ReliefF method is as follows:

$$\begin{aligned}
CF(RF) = &(0.00532) \times R_b + (0.02866) \times D_d + (0.05694) \times F_u \\
&+ (0.04566) \times R_t + (0.02452) \times L_o + (-0.03597) \\
&\times R_f + (-0.03246) \times B_s + (-0.03282) \times R_e \\
&+ (-0.0182) \times C_c + (0.05803) \times I_f + (-0.0198) \times R_c \\
&+ (0.02878) \times C + (0.01382) \times R_h + (0.05113) \times R_n \\
&+ (0.02103) \times STI + (0.02622) \times SPI + (0.03788) \\
&\times TWI
\end{aligned} \tag{13}$$

As observed in Table 7 and Eq. 13, the highest CF value is 14.37, corresponding to watershed 51, and the second-highest value is 12.78, corresponding to watershed 50; likewise, other subsequent CF values are calculated, which are shown in Table 7 and Fig. 4b.

Pearson correlation coefficient (PCC)

The Pearson correlation coefficient (C_r) is generally used to measure and rank the relationship between the calculated and predicted values. A value near $+1$ indicates a perfect correlation from the model output, whereas the value of -1 indicates negative correlations. If a value is closer to 0, it means that there is no direct correlation between the calculated and predicted parameters (Gibbons 2014).

$$\overline{X_p} = \sum_{i=1}^{m} \frac{X_{(p,i)}}{m}; \overline{Y_q} = \sum_{j=1}^{m} \frac{Y_{(q,j)}}{m} \tag{14}$$

$$PCC_{(p,q)} = \frac{\sum_{i=1}^{m} \left(X_{(p,i)} - \overline{X_p} \right) \left(Y_{(q,i)} - \overline{Y_q} \right)}{\sqrt{\sum_{i=1}^{m} \left(X_{(p,i)} - \overline{X_p} \right)^2 \sum_{j=1}^{m} \left(Y_{(q,j)} - \overline{Y_q} \right)^2}} \tag{15}$$

Here, $PCC_{(p,q)}$ gives the value of the Pearson linear correlation coefficient, m is the length of parameters, and $\overline{X_p}$ and $\overline{Y_q}$ give the values of a mean of each parameter.

The Pearson correlation coefficient method was used to assign the ranks to the morphometric parameters. I_f is assigned rank 1 with a weightage of 0.93, whereas F_u is assigned rank 2 with a weightage of 0.9197. Other subsequent values, ranks, and weights are cited in Table 6. The equation of compound factor based on the weightage of correlation coefficient ranked features is as follows:

$$\begin{aligned}
CF(PCC) = &(0.3689) \times R_b + (0.8069) \times D_d + (0.9197) \times F_u \\
&+ (0.8659) \times R_t + (-0.7855) \times L_o + (0.1328) \times R_f \\
&+ (-0.071) \times B_s + (0.1487) \times R_e + (-0.0267) \times C_c \\
&+ (0.9373) \times I_f + (0.0484) \times R_c + (0.8161) \times C \\
&+ (0.2701) \times R_h + (0.9104) \times R_n + (0.7203) \times STI \\
&+ (0.7744) \times SPI + (0.8488) \times TWI
\end{aligned} \tag{16}$$

Table 4 Preliminary priority ranking of linear, areal, and shape parameters

WS no	R_b	D_d	F_u	R_t	L_o	R_f	B_s	R_e	C_c	I_f	R_c	C	R_h	R_n	STI	SPI	TWI
1	8	4	7	6	48	11	38	11	32	7	19	4	6	3	3	5	7
2	37	20	27	28	31	24	26	26	43	24	9	19	2	8	7	8	8
3	21	18	10	1	32	1	50	2	18	12	34	18	4	2	2	2	4
4	14	1	4	18	51	44	7	44	48	2	4	1	8	4	5	6	5
5	28	5	2	5	45	29	21	27	37	3	14	5	1	1	1	1	1
6	30	33	19	14	17	17	33	17	9	22	42	33	10	11	14	10	13
7	29	7	5	9	46	30	16	32	27	5	24	6	11	6	8	7	6
8	1	12	3	2	40	38	13	33	2	4	50	12	7	7	6	4	3
9	24	9	6	19	41	50	2	50	12	6	40	8	5	18	11	12	11
10	9	30	9	21	22	49	3	49	10	14	43	30	3	26	9	9	10
11	32	3	1	3	49	25	24	28	42	1	10	3	17	5	4	3	2
12	4	14	11	11	36	26	28	23	40	10	11	14	32	16	12	11	9
13	36	50	47	39	2	2	45	3	41	49	12	50	26	38	19	22	17
14	13	13	8	7	39	12	40	12	29	8	20	13	40	12	20	14	16
15	31	25	17	15	26	3	47	4	49	19	2	24	22	9	17	16	14
16	16	26	26	22	27	4	51	1	33	26	21	26	45	17	27	25	20
17	5	10	12	16	42	18	31	18	47	9	5	10	25	10	18	15	12
18	34	32	28	26	20	27	27	24	30	28	22	32	23	20	42	30	31
19	33	2	20	23	50	39	11	40	4	11	48	2	44	14	30	21	15
20	46	41	30	30	11	31	22	29	23	35	28	40	13	27	35	28	22
21	15	15	18	17	37	13	37	13	38	17	15	15	33	13	37	35	23
22	42	17	49	49	35	32	17	34	24	44	29	17	39	28	16	24	18
23	25	21	16	8	33	5	48	5	22	16	30	21	46	15	34	33	30
24	11	31	35	27	21	7	46	6	25	33	27	31	42	19	40	29	33
25	41	43	29	25	8	19	32	19	15	34	35	43	19	29	43	43	28
26	48	35	25	31	18	47	5	47	14	27	38	35	18	45	33	31	21
27	35	48	32	29	4	20	34	20	35	37	16	47	16	33	25	19	29
28	17	47	48	50	5	40	12	41	39	50	13	48	12	46	47	37	32
29	43	37	50	48	15	14	39	14	44	48	8	37	36	34	13	26	34
30	20	29	24	24	23	15	36	15	34	25	17	29	38	23	29	18	24
31	18	40	41	37	12	33	18	35	11	42	41	41	20	35	44	32	43
32	49	38	40	44	14	41	14	36	45	38	7	38	24	39	23	40	40
33	38	42	39	36	10	23	29	25	26	41	25	42	27	36	41	23	37
34	2	36	21	12	16	6	49	7	31	23	23	36	48	24	31	27	25
35	26	16	14	10	38	34	19	37	6	13	44	16	43	30	46	34	36
36	22	39	42	32	13	8	42	8	17	43	36	39	47	47	39	38	41
37	10	46	43	35	6	9	43	9	19	46	32	46	29	40	48	36	47
38	40	34	36	45	19	35	20	38	50	36	3	34	31	41	36	49	49
39	47	24	33	38	28	46	6	46	20	32	33	25	28	42	21	20	26
40	39	6	38	40	47	45	8	45	13	29	39	7	50	43	22	50	46
41	7	27	13	4	24	10	44	10	7	15	45	27	49	22	38	41	27
42	50	8	22	33	43	51	1	51	5	20	47	9	21	37	10	17	39
43	23	44	37	34	9	21	35	21	36	40	18	44	30	44	28	45	42
44	12	11	23	20	44	22	30	22	1	21	51	11	51	25	32	42	38
45	6	22	15	13	29	36	23	30	8	18	46	22	35	31	45	47	35
46	27	19	34	41	34	37	15	39	46	31	6	20	34	32	26	48	48
47	51	23	31	42	30	48	4	48	21	30	31	23	41	49	15	51	44
48	3	49	44	46	3	42	9	42	16	47	37	49	15	50	51	46	45
49	44	51	51	51	1	28	25	31	28	51	26	51	9	51	49	44	51
50	45	28	45	47	25	16	41	16	51	39	1	28	37	21	50	39	50
51	19	45	46	43	7	43	10	43	3	45	49	45	14	48	24	13	19

Table 5 Weight assigned by WSA techniques

| Parameters	R_b	D_d	F_u	R_t	L_o	R_f	B_s	R_e	C_c	I_f	R_c	C	R_h	R_n	STI	SPI	TWI
R_b	1	0.143167	0.395837	0.511403	-0.13683	0.267602	-0.2705	0.298462	0.148507	0.351041	-0.15357	0.142805	-0.04688	0.327692	-0.02462	0.181086	0.264253
D_d	0.143167	1	0.734118	0.581538	-0.99765	-0.16371	0.208778	-0.15837	0.03819	0.844163	-0.02326	0.999457	-0.03348	0.675566	0.590679	0.433484	0.502805
F_u	0.395837	0.734118	1	0.919367	-0.71719	0.011765	0.022443	0.035294	0.104615	0.979095	-0.08299	0.743077	0.232036	0.784072	0.551041	0.594299	0.712217
R_t	0.511403	0.581538	0.919367	1	-0.56443	0.32543	-0.29502	0.351222	0.16371	0.877466	-0.14145	0.591403	0.095023	0.797285	0.40914	0.553846	0.679819
L_o	-0.13683	-0.99765	-0.71719	-0.56443	1	0.167149	-0.21077	0.16371	-0.039	-0.83077	0.024706	-0.99674	0.067511	-0.65466	-0.56706	-0.40081	-0.47158
R_f	0.267602	-0.16371	0.011765	0.32543	0.167149	1	-0.98724	0.991493	-0.29864	-0.02986	0.304163	-0.1552	-0.2857	0.355385	-0.09765	0.086063	0.099548
B_s	-0.2705	0.208778	0.022443	-0.29502	-0.21077	-0.98724	1	-0.98751	0.296833	0.065882	-0.299	0.200905	0.34009	-0.30199	0.16905	-0.00606	-0.01738
R_e	0.298462	-0.15837	0.035294	0.351222	0.16371	0.991493	-0.98751	1	-0.28262	-0.00977	0.287421	-0.14968	-0.28688	0.369321	-0.09176	0.087149	0.113756
C_c	0.148507	0.03819	0.104615	0.16371	-0.039	-0.29864	0.296833	-0.28262	1	0.084615	-0.99747	0.033213	-0.07557	-0.19041	-0.14271	-0.06045	-0.04217
I_f	0.351041	0.844163	0.979095	0.877466	-0.83077	-0.02986	0.065882	-0.00977	0.084615	1	-0.0638	0.851222	0.153846	0.80371	0.577376	0.580543	0.692217
R_c	-0.15357	-0.02326	-0.08299	-0.14145	0.024706	0.304163	-0.299	0.287421	-0.99747	-0.0638	1	-0.01783	0.085249	0.213937	0.153303	0.082986	0.059367
C	0.142805	0.999457	0.743077	0.591403	-0.99674	-0.1552	0.200905	-0.14968	0.033213	0.851222	-0.01783	1	-0.0248	0.687692	0.596742	0.443801	0.515837
R_h	-0.04688	-0.03348	0.232036	0.095023	0.067511	-0.2857	0.34009	-0.28688	-0.07557	0.153846	0.085249	-0.0248	1	0.23086	0.41276	0.559186	0.478552
R_n	0.327692	0.675566	0.784072	0.797285	-0.65466	0.355385	-0.30199	0.369321	-0.19041	0.80371	0.213937	0.687692	0.23086	1	0.545973	0.697647	0.765611
STI	-0.02462	0.590679	0.551041	0.40914	-0.56706	-0.09765	0.16905	-0.09176	-0.14271	0.577376	0.153303	0.596742	0.41276	0.545973	1	0.750136	0.732127
SPI	0.181086	0.433484	0.594299	0.553846	-0.40081	0.086063	-0.00606	0.087149	-0.06045	0.580543	0.082986	0.443801	0.559186	0.697647	0.750136	1	0.885339
TWI	0.264253	0.502805	0.712217	0.679819	-0.47158	0.099548	-0.01738	0.113756	-0.04217	0.692217	0.059367	0.515837	0.478552	0.765611	0.732127	0.885339	1
Summation of correlation	3.399457	5.375475	7.019095	6.855747	-5.16443	1.590588	-1.07149	1.731222	-0.25937	6.926968	0.431765	5.4619	2.90181	7.107692	5.564525	6.468235	6.970317
Total	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095	61.3095
Weightage	0.055447	0.087678	0.114486	0.111822	-0.08424	0.025944	-0.01748	0.028237	-0.00423	0.112984	0.007042	0.089087	0.047331	0.115931	0.090761	0.105501	0.113691

Fig. 4 Watershed prioritized maps using **a** CF (WSA), **b** CF (RF), and **c** CF (PCC)
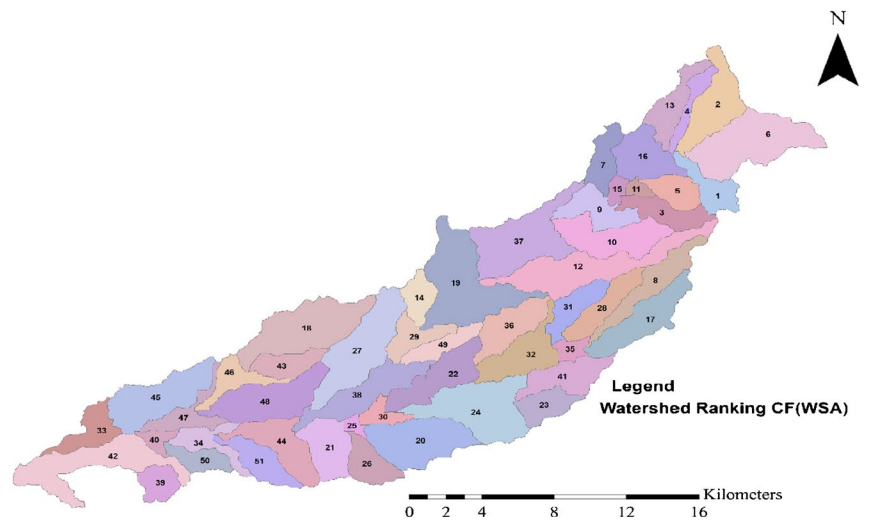
(a)
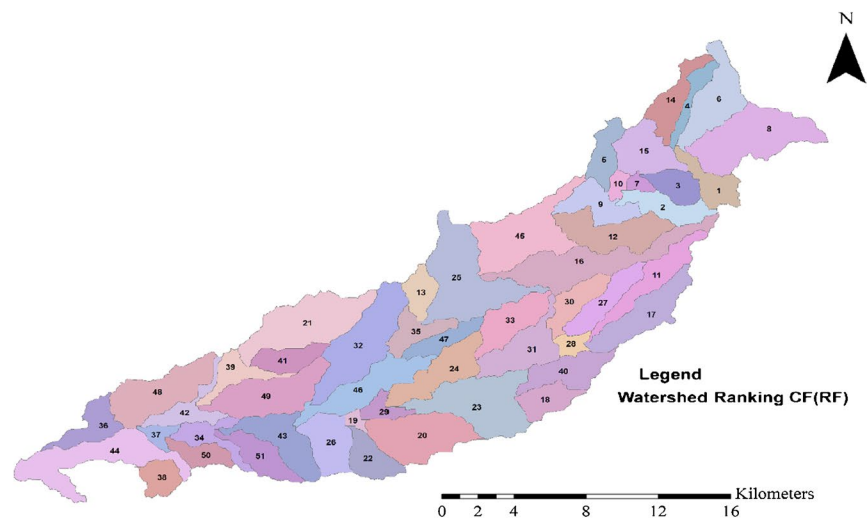
(b)
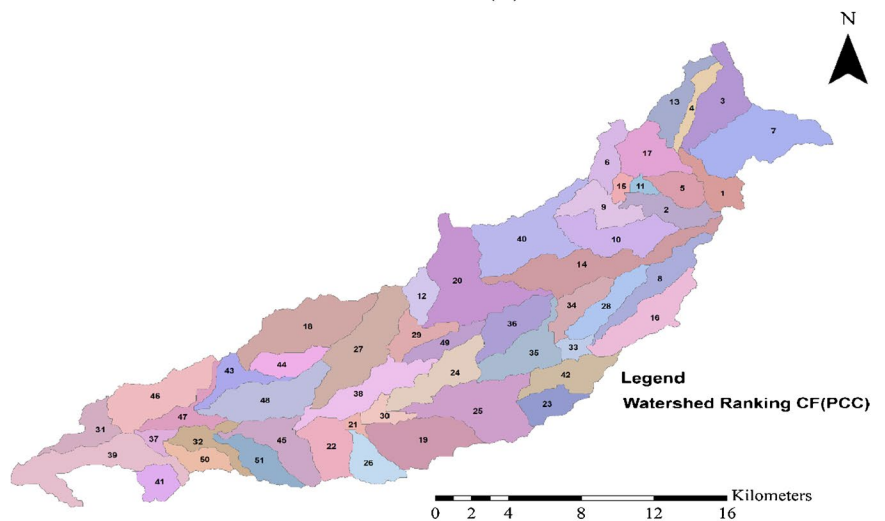
(c)

Table 6 Weightage of parameters based on WSA, ReliefF, and Pearson correlation coefficient

Sr. No	Parameters	Weightage WSA	Weightage ReliefF	Weightage Pearson correlation coefficient
1	Rb	0.0554	0.00532	0.3689
2	Dd	0.0876	0.02866	0.8069
3	Fu	0.1144	0.05694	0.9197
4	Rt	0.111822	0.04566	0.8659
5	Lo	−0.08424	0.02452	−0.7855
6	Rf	0.025944	−0.03597	0.1328
7	Bs	−0.01748	−0.03246	−0.071
8	Re	0.028237	−0.03282	0.1487
9	Cc	−0.00423	−0.0182	−0.0267
10	If	0.112984	0.05803	0.9373
11	Rc	0.007042	−0.0198	0.0484
12	C	0.089087	0.02878	0.8161
13	Rh	0.047331	0.01382	0.2701
14	Rn	0.115931	0.05113	0.9104
15	STI	0.090761	0.02103	0.7203
16	SPI	0.105501	0.02622	0.7744
17	TWI	0.113691	0.03788	0.8488

As observed in Table 7 and Eq. 16, the highest CF value is 405.67, and it is for watershed 51, the second-highest value is 375.3, and it is for watershed 49. The watershed map constructed is shown in Fig. 4c.

Results and discussion

The present paper calculates and derives the basic morphometric parameters using RS and GIS techniques. Various parameters are mentioned in Tables 2 and 3, and the ranks are assigned based on the relation of parameters with soil erodibility, which is shown in Table 4. In the present study, the prioritization of the watershed was decided based on CF values calculated through WSA, ReliefF, and the Pearson correlation coefficient method. Validation and comparison of CF values have been done using MLP and SVM models.

Watershed priority based on CF values

The watershed CF values are calculated using WSA, ReliefF, and the correlation coefficient method. For prioritization, the lowest value of CF is given the priority rank 1, the next lower value is given the priority rank of 2, and so on for all the 51 microwatersheds; CF values are calculated through the three methods mentioned above. Furthermore, the prioritized watersheds are categorized into five categories, i.e., very high, high,

moderate, low, and very low. Finally, the categories of watershed maps are utilized for the decision-making system.

As observed in Table 7, the lowest CF values based on WSA are 1.34, hence, it is assigned rank 1, and the same ranking is assigned, which will be useful for prioritizing the watersheds (Table 7). The CF values calculated and the prioritized watersheds based on WSA, ReliefF, and Pearson correlation coefficient are shown. For CF values obtained through ReliefF, all the 51 microwatersheds of the Rel River catchment are classified into five priority categories such as (i) very high (− 1.355 to ≤ 1.789), (ii) high (1.789 to ≤ 4.933), (iii) medium (4.933 to ≤ 8.078), (iv) low (8.078 to ≤ 11.223), and (v) very low (> 11.223) as given in Table 8. It is observed through Table 8 that the 8 microwatersheds belong to very high category (WS no. 1, 3, 4, 5, 7, 8, 9, and 11), 8 microwatersheds under a high category (WS no. 2, 6, 10, 12, 14, 15, 17, and 19), 14 microwatersheds under medium (WS no. 16, 18, 20, 21, 23, 26, 30, 34, 35, 39, 41, 42, 44, and 45), 12 microwatersheds under a low category (WS no. 22, 24, 25, 27, 31, 32, 33, 38, 40, 46, 47, and 51), and 9 microwatershed under the very low category (WS no. 13, 28, 29, 36, 37, 43, 48, 49, and 50). The CF values obtained from Eqs. 11, 13, and 16 watershed are constructed, shown in Fig. 4a–c.

ML model predictions

Two ML algorithms, MLP and SVM, have been utilized to predict the CF values for the prioritization of watersheds. The training and Tenfold cross-validation are performed on all the

Table 7 Final priority ranking based on CF (WSA), CF (RF), and CF (CC)

Watershed no	CF (WSA)	Rank (WSA)	CF (RF)	Rank (RF)	CF (PCC)	Rank (PCC)
1	1.528264	2	0.27492	6	6.7236	3
2	16.01509	13	3.86089	14	122.861	13
3	4.903464	6	0.91985	8	40.4664	7
4	3.938099	4	−0.89169	4	16.962	4
5	1.335835	1	−1.35562	1	1.6663	1
6	18.4453	16	3.9834	15	147.303	17
7	5.845198	7	0.06385	5	34.9447	6
8	4.310874	5	−0.92929	3	26.2914	5
9	11.61175	11	0.61581	7	77.8723	11
10	17.83164	15	2.26766	10	136.1282	15
11	1.943757	3	−1.03453	2	3.8269	2
12	10.66848	9	2.01432	9	77.2943	9
13	36.81942	37	11.72068	45	299.5558	40
14	10.70105	10	2.52398	12	77.4664	10
15	15.71252	12	4.34046	16	124.3031	14
16	22.29458	19	6.8578	25	174.8333	20
17	9.879781	8	2.45127	11	70.8866	8
18	29.97648	28	7.48339	27	232.9969	28
19	16.69544	14	3.32148	13	112.5607	12
20	33.24179	31	8.05952	30	261.6969	34
21	18.98612	17	5.2402	17	142.7939	16
22	31.16441	29	9.42732	35	234.603	29
23	20.29711	18	5.57612	21	154.4773	18
24	29.3392	27	8.96685	32	231.492	27
25	35.69458	36	9.17599	33	282.7589	36
26	34.10074	35	7.56216	28	261.2875	33
27	33.69377	32	8.90357	31	270.836	35
28	45.27464	49	12.40651	47	358.6792	49
29	37.55711	38	11.83673	46	296.0889	38
30	24.4378	22	6.78556	24	190.649	24
31	39.75809	43	10.70476	41	312.3009	44
32	40.25462	46	10.58357	39	311.6823	43
33	38.47929	41	10.62681	40	303.8441	42
34	24.55777	24	6.6908	23	195.7769	25
35	24.50611	23	5.31351	18	179.0141	23
36	40.02803	45	12.51027	48	316.2031	46
37	41.60316	48	12.74387	49	333.7207	48
38	41.5177	47	11.12328	42	319.6869	47
39	31.92323	30	7.92383	29	240.4292	30
40	33.82409	33	9.44803	36	244.6926	31
41	22.54344	20	5.53433	20	174.0157	19
42	24.59601	25	5.48176	19	175.2774	21
43	39.93747	44	11.28255	43	315.9518	45
44	24.13611	21	6.89754	26	176.7408	22
45	26.92503	26	5.98962	22	202.8793	26
46	33.90714	34	9.38026	34	254.4021	32
47	38.02727	40	9.71542	37	283.0729	37
48	47.48223	50	12.77914	50	375.3013	50
49	51.12991	51	14.36773	51	405.6693	51
50	38.61711	42	11.60586	44	299.0939	39
51	38.00189	39	9.95714	38	300.9122	41

Table 8 Priority category based on CF (RF) values

Zone	CF(RF) value	Watershed number
Very high	−1.355 to 1.789	1, 3, 4, 5, 7, 8, 9, and 11
High	1.789 to 4.933	2, 6, 10, 12, 14, 15, 17, and 19
Moderate	4.933 to 8.078	16, 18, 20, 21, 23, 26, 30, 34, 35, 39, 41, 42, 44, and 45
Low	8.078 to 11.223	22, 24, 25, 27, 31, 32, 33, 38, 40, 46, 47, and 51
Very low	>11.223	13, 28, 29, 36, 37, 43, 48, 49, and 50

51 microwatersheds, and the derived parameters are normalized in the range [−1 to +1] to minimize the biasing error. Results of the prediction are shown in Fig. 5a–f. The CF (WSA) values predicted through SVM and MLP are shown in Fig. 5a, b. It is observed that SVM accurately predicts the CF values with a very less average error of 0.04 whereas there is a slight deviation in prediction results with MLP with an average error of 0.90 for all the watershed values considered. Similarly when CF (RF) is predicted, then, an average error of 0.02 and 0.28 is observed from SVM and MLP, respectively, as shown in Fig. 5c, d. Further, the average error of 0.36 and 7.65 is observed when CF (PCC) is predicted, as shown in Fig. 5e, f. The results can infer that the best CF values for watershed prioritization should be obtained from

Fig. 5 a–f Tenfold cross validation prediction results using SVM and ANN
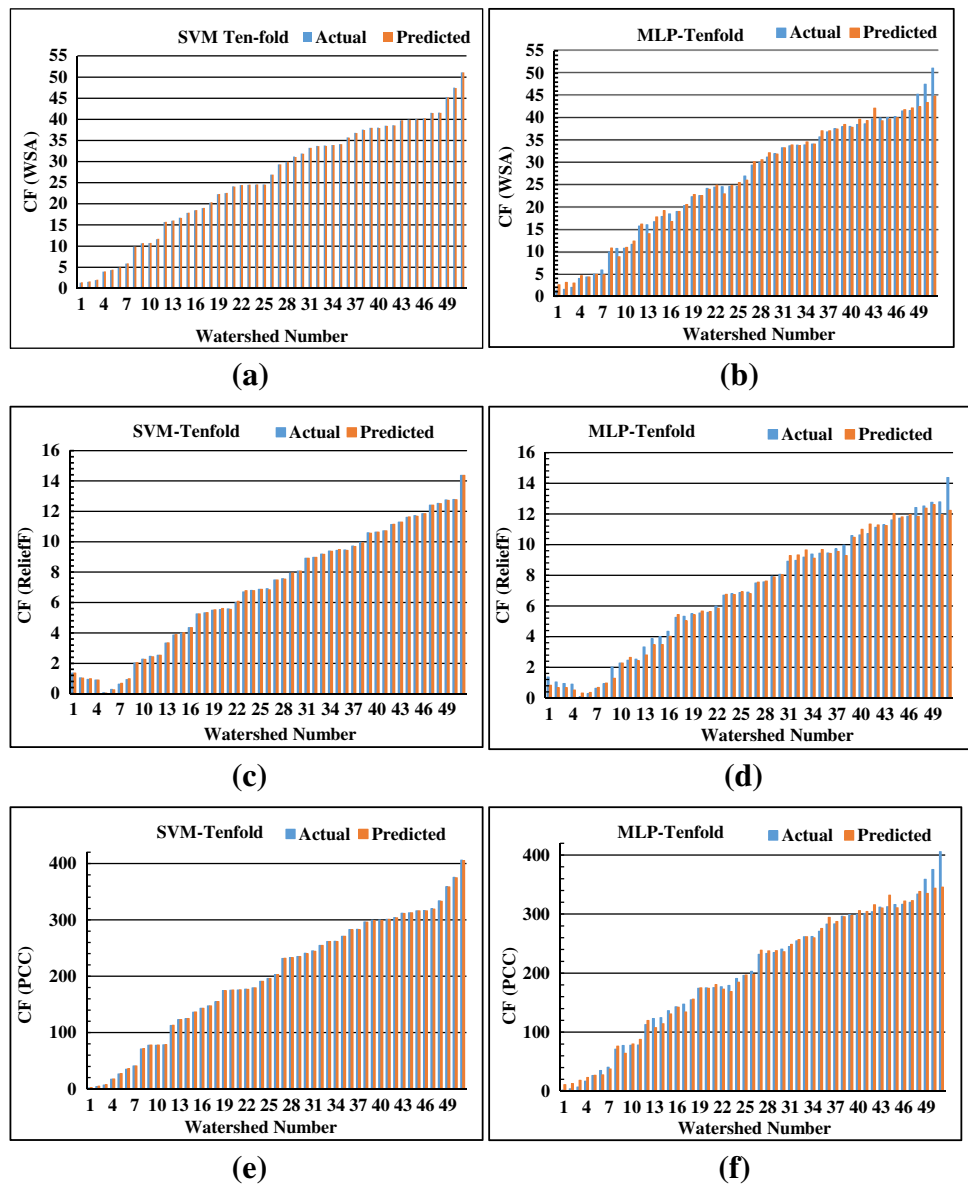
ReliefF Eq. 13 as there are very few mean errors observed from both SVM and MLP. Moreover, to evaluate the suitability of ML algorithms, three performance metrics: mean absolute error, correlation coefficient, and root mean square error, are calculated, and the results are discussed below.

Correlation coefficient

The correlation coefficient (R) is calculated to find the relationship between the dependent and independent variables. When R is observed as $+1$, it indicates a perfect correlation between the two variables, while -1 indicates a negative correlation between two variables. The η shows the total number of observations, $\sum a$ shows the mean value of all the first variables in the data set, and $\sum b$ represents the mean value of all the second variables in the data set. It is mathematically calculated by

$$R = \frac{\eta \sum ab - \sum a \sum b}{\sqrt{\eta \sum a^2 - \sum a^2 [\eta \sum b^2 - \left(\sum b\right)^2}} \tag{17}$$

Mean absolute error

Mean absolute error (MAE) has been used to analyze the performance of the ML model. The difference between the predicted value and the actual value is calculated by

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \odot_i - \odot \right| \tag{18}$$

Here, $\odot_i$ shows the predicted value by the machine learning model, and $\odot$ shows the actual value from experiments.

Root mean square error

Root mean square error (RMSE) represents the standard deviation of the residuals. RMSE is commonly used in hydrology, forecasting, and regression analysis to verify experimental results.

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} \left(\odot_i - \odot \right)^2} \tag{19}$$

where i is variable, n represents data points, $\odot i$ represent actual observations, and $\odot$ represents predicted observation.

Tables 9 and 10 show the results of SVM Tenfold and MLP Tenfold for CF (WSA), CF (RF), and CF (CC) values. The prediction of ML models is compared with three performance metrics correlation coefficient, MAE, and RMSE values. Initially, the performance of the ML model is evaluated through a correlation coefficient. It is observed that there is no significant deviation in performance metrics from both the SVM and MLP

Table 9 Statistical evaluation of SVM Tenfold model

Parameters	Ideal values	CF (WSA)	CF (RF)	CF (CC)
Correlation coefficient	1	1	1	1
MAE	0	0.0472	0.0236	0.3684
RMSE	0	0.0565	0.032	0.4361

Table 10 Statistical evaluation of MLP Tenfold model

Parameters	Ideal values	CF (WSA)	CF (RF)	CF (CC)
Correlation coefficient	1	0.9946	0.9949	0.9938
MAE	0	0.9021	0.2856	7.6514
RMSE	0	1.4233	0.4402	12.2689

models. The correlation coefficient was observed as 1, which signifies a very good correlation between the calculated and predicted values.

Figure 6a, b shows the SVM Tenfold and MLP Tenfold results to compare the prediction of CF values using MAE and RMSE. From Fig. 6, it is clear that the minimum MAE is 0.0236 for CF (RF) using the SVM Tenfold model compared to the minimum MAE of 0.2856 for CF (RF) using the MLP Tenfold model. Furthermore, the minimum RMSE is 0.032 for CF (RF) using the SVM Tenfold model compared to the minimum RMSE of 0.440 for CF (RF) using the MLP Tenfold model. The maximum MAE is 0.3684 for CF (CC) using an SVM Tenfold compared to the maximum MAE of 7.6514 for CF (CC) using MLP Tenfold. Also, the maximum RMSE observed is 0.4361 for CF (CC) using SVM Tenfold, compared to the maximum RMSE of 12.26 for CF (RF) using the MLP Tenfold model. Results show that the SVM Tenfold gives better prediction capability as compared to MLP Tenfold for the prediction of CF values with CF (RF), followed by CF (WSA) and CF (PCC), respectively.

Figure 7a–c shows the variation in MAE and RMSE for the CF (RF), CF (WSA), and CF (PCC) predicted using the SVM Tenfold and MLP Tenfold models. The minimum MAE observed is 0.0236 with CF (RF), whereas the minimum MAE is 0.3684 for CF (PCC) from SVM Tenfold. The maximum MAE is 0.2856 with CF (RF) compared to 7.6514 with CF (CC) from MLP Tenfold. Furthermore, the minimum RMSE for predicting CF (RF) is 0.032 compared to 0.4361 in CF (CC) using SVM Tenfold. The maximum RMSE for predicting CF (RF) is 0.4402 compared to 12.2689 in CF (CC) using MLP Tenfold. It is observed that CF values calculated from ReliefF exhibit a higher correlation coefficient as well as very low MAE and RMSE values with both SVM and ANN models (Tables 9 and 10); therefore, CF values calculated from ReliefF are used for constructing watershed prioritization. The final watershed priority category map of 51 microwatersheds is shown in

Fig. 6 Tenfold cross validation prediction results using **a** SVM and **b** ANN
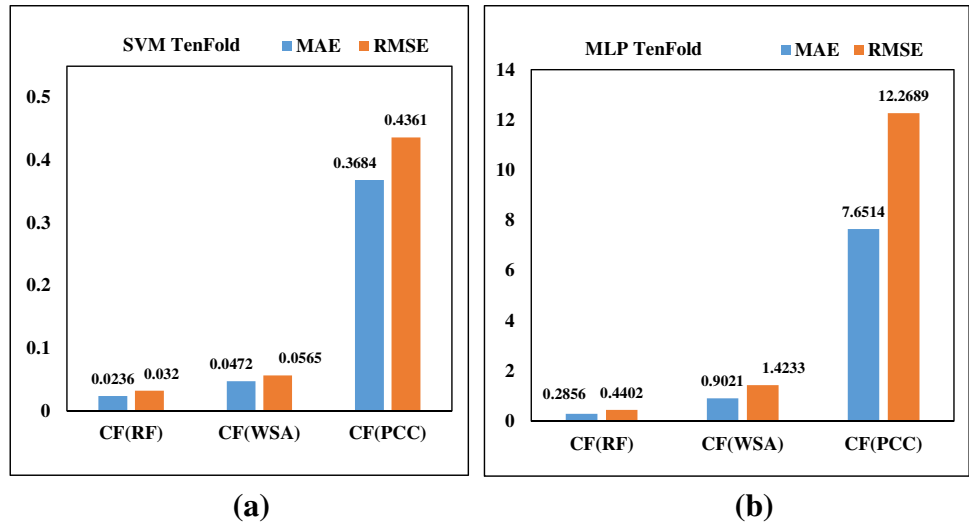

(a)

(b)

Fig. 8. It is observed that the percentage area of microwatersheds under the very high category is 13.45%, for a high category, it is 15.72%, for medium category is 30.24%, low category is 18.81%, and for very low category, it is 21.78%.

This information is beneficial in implementing water management strategies regarding soil and water conservation measures. The result also shows its vulnerability to erosion and runoff potential. This watershed is mainly situated in

Fig. 7 MAE and RMSE variations with SVM and MLP for **a** CF (RF) and **b** CF (WSA) and CF (CC)

(a)

(b)

(c)
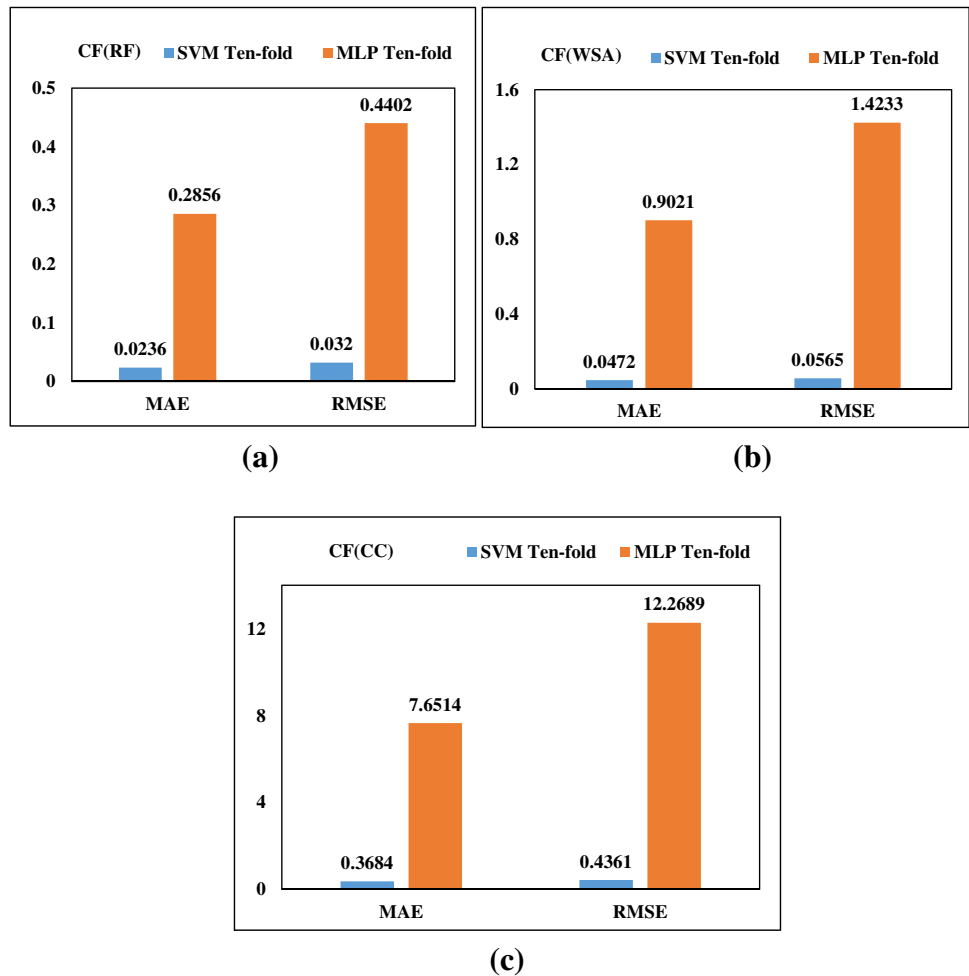
Fig. 8 Watershed priority map based on CF (RF) values
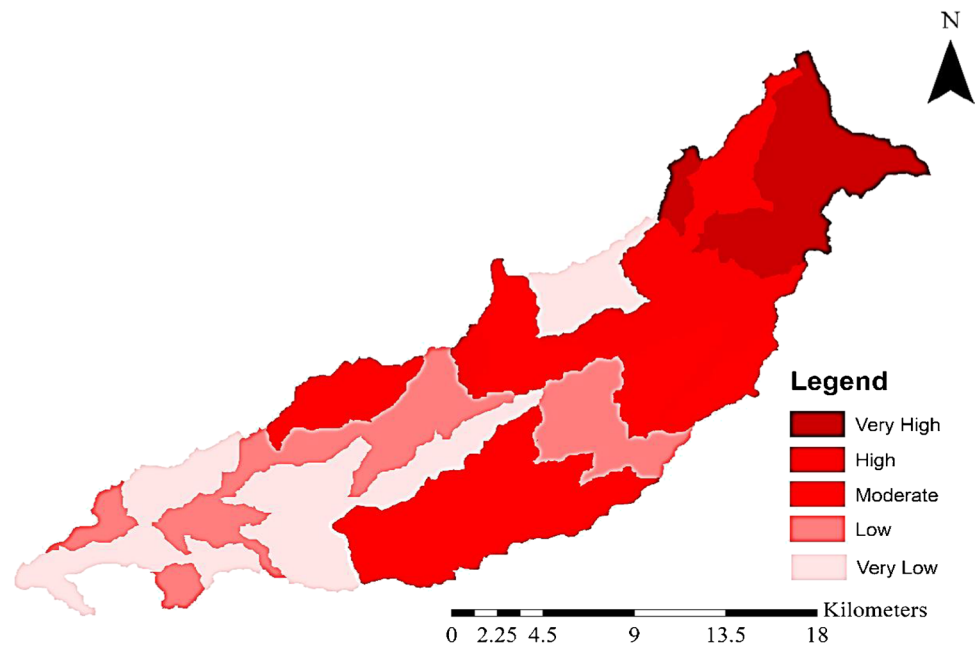
the upstream part of the catchment or basin and high slopes and elongated basins.

Conclusion

In the methodology proposed, the utility of feature ranking is initially explored in detail for estimating the weightage of various morphological and topological parameters. Afterward, the CF values for various watersheds are calculated from WSA, ReliefF, and Pearson correlation coefficient, obtained from the assigned weightage of various morphological and topological parameters. To determine the best CF values, ML algorithms are explored in detail. Finally, the performance of models is estimated after comparing the three parameters, i.e., mean absolute error, correlation coefficient, and root mean square error. The findings are mentioned below:

1. The SVM model gives better results for predicting various watersheds as compared to MLP model.
2. The comparison of MAE and RMSE for predicting CF (RF), CF (WSA), and CF (CC) reveals that CF (RF) is the best model for the prediction of CF values as compared to other calculations; hence, it is utilized for prioritization of watersheds.
3. It is suggested that watersheds 1, 3, 4, 5, 7, 8, 9, and 11 have a very high vulnerability to soil erosion, followed by the rest of the watersheds in the study region. Hence, appropriate soil and water conservation measures should be adopted to protect against degradation.

Based on a study conducted, the integrated framework of RS, GIS, feature ranking, and machine learning seems to be an efficient water resource management technique for watershed ranking and prioritization. Tenfold cross-validation and feature ranking is a novel approach for accurately predicting CF for watershed ranking and prioritization. The proposed methodology should be extended to investigate the influence of more RS and GIS techniques as well as exploration of more morphometric and topological parameters and analyze its effect on constructing various watersheds for soil and water conservation. It should be noted that the performance of ML models for prediction is dependent on the calculated or extracted parameters.

Data availability The data required for watershed parameter delineation and prioritization were supported by SAC-ISRO.

Code availability MATLAB code is utilized for the preparation of ML models.

Declarations

Competing interests The authors declare that they have no competing interests.

References

Aher PD, Adinarayana J, Gorantiwar SD (2014) Quantification of morphometric characterization and prioritization for management planning in semi-arid tropics of India: a remote sensing and GIS approach. J Hydrol 511:850–860. https://doi.org/10.1016/J.JHYDROL.2014.02.028

Akhenia P, Bhavsar K, Panchal J, Vakharia V (2021) Fault severity classification of ball bearing using SinGAN and deep convolutional neural network. Proc Inst Mech Eng Part C J Mech Eng Sci 095440622110431. https://doi.org/10.1177/09544062211043132

Alilou H, Rahmati O, Singh VP et al (2019) Evaluation of watershed health using Fuzzy-ANP approach considering geo-environmental and topo-hydrological criteria. J Environ Manage 232:22–36. https://doi.org/10.1016/J.JENVMAN.2018.11.019

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci Bull 24:43–69. https://doi.org/10.1080/02626667909491834

Chowdary VM, Chakraborthy D, Jeyaram A et al (2013) Multi-criteria decision making approach for watershed prioritization using analytic hierarchy process technique and GIS. Water Resour Manag 27:3555–3571. https://doi.org/10.1007/S11269-013-0364-6

Darji K, Patel D, Dubey AK, et al (2021) An approach of satellite and UAS based mosaicked DEM for hydrodynamic modelling -a case of flood assessment of Dhanera City, Gujarat, India. J geomatics 247–257

Dave V, Singh S, Vakharia V (2020) Diagnosis of bearing faults using multi fusion signal processing techniques and mutual information. Indian J Eng Mater Sci 27:878–888

Faniran A (1968) The index of drainage intensity: a provisional new drainage factor. Aust J Sci 31:326–330

Farhan Y, Anbar A, Al-Shaikh N, Mousa R (2017) Prioritization of semi-arid agricultural watershed using morphometric and principal component analysis, remote sensing, and GIS techniques, the Zerqa River Watershed, Northern Jordan. Agric Sci 08:113–148. https://doi.org/10.4236/AS.2017.81009

Gibbons JD CS (2014) Nonparametric statistical inference, Fourth Edition. Nonparametric Stat Inference, Fourth Ed. https://doi.org/10.4324/9780203911563

Horton RE (1932) Drainage-basin characteristics. Eos, Trans Am Geophys Union 13:350–361. https://doi.org/10.1029/TR013I001P00350

Horton RE (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Geol Soc Am Bull 56(3):275–370

Jaiswal RK, Ghosh NC, Galkate RV, Thomas T (2015) Multi criteria decision analysis (MCDA) for watershed prioritization. Aquat Procedia 4:1553–1560. https://doi.org/10.1016/J.AQPRO.2015.02.201

Kadam AK, Jaweed TH, Umrikar BN et al (2017) Morphometric prioritization of semi-arid watershed for plant growth potential using GIS technique. Model Earth Syst Environ 3:1663–1673. https://doi.org/10.1007/S40808-017-0386-9

Kankar PK, Sharma SC, Harsha SP (2011) Rolling element bearing fault diagnosis using autocorrelation and continuous wavelet transform. Jvc/J Vib Control 17:2081–2094. https://doi.org/10.1177/1077546310395970

Khan MA, Gupta VP, Moharana PC (2001) Watershed prioritization using remote sensing and geographical information system: a case study from Guhiya, India. J Arid Environ 49:465–475. https://doi.org/10.1006/JARE.2001.0797

Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 784 LNCS:171–182. https://doi.org/10.1007/3-540-57868-4_57

Kulimushi LC, Bashagaluke JB, Choudhari P, et al (2021a) Novel combination of analytical hierarchy process and weighted sum analysis for watersheds prioritization. A study of Ulindi catchment, Congo River Basin. Geocarto Int 1–39. https://doi.org/10.1080/10106049.2021.2002426

Kulimushi LC, Choudhari P, Maniragaba A et al (2021b) Erosion risk assessment through prioritization of sub-watersheds in Nyabarongo river catchment. Rwanda Environ Challenges 5:100260. https://doi.org/10.1016/j.envc.2021.100260

Malik A, Kumar A, Kandpal H (2019) Morphometric analysis and prioritization of sub-watersheds in a hilly watershed using weighted sum approach. Arab J Geosci 12https://doi.org/10.1007/S12517-019-4310-7

Melton MA (1957) An analysis of the relations among elements of climate, surface properties, and geomorphology. Columbia Univ New York

Memon N, Patel DP, Bhatt N, Patel SB (2020) Integrated framework for flood relief package (FRP) allocation in semiarid region: a case of Rel River flood, Gujarat, India. Nat Hazards 100:279–311. https://doi.org/10.1007/S11069-019-03812-Z

Meshram SG, Sharma SK (2017) Prioritization of watershed through morphometric parameters: a PCA-based approach. Appl Water Sci 7:1505–1519. https://doi.org/10.1007/S13201-015-0332-9

Miller VC (1953) A quantitative geomorphic study of drainage basin characteristics in the clinch mountain area Virginia and Tennessee. Columbia Univ New York

Moore ID, Burch GJ (1986) Sediment transport capacity of sheet and rill flow: application of unit stream power theory. Water Resour Res 22:1350–1360. https://doi.org/10.1029/WR022I008P01350

Naseriparsa M, Bidgoli A-M, Varaee T (2014) A hybrid feature selection method to improve performance of a group of classification algorithms. Int J Comput Appl 69:975–8887. https://doi.org/10.5120/12065-8172

Nooka Ratnam K, Srivastava YK, Venkateswara Rao V et al (2005) Check Dam positioning by prioritization micro-watersheds using SYI model and morphometric analysis - remote sensing and GIS perspective. J Indian Soc Remote Sens 33:25–38. https://doi.org/10.1007/BF02989988

Patel A, Singh MM, Singh SK et al (2022) AHP and TOPSIS based sub-watershed prioritization and tectonic analysis of Ami River Basin, Uttar Pradesh. J Geol Soc India 98:423–430. https://doi.org/10.1007/s12594-022-1995-0

Patel D, Srivastava P, Gupta M (2015) Decision support system integrated with geographic information system to target restoration actions in watersheds of arid environment: a case study of Hathmati. J Earth Syst Sci 124:71–86. https://doi.org/10.1007/s12040-014-0515-z

Patel DR, Kiran MB, Vakharia V (2020) Modeling and prediction of surface roughness using multiple regressions: A noncontact approach. Eng R 2(2):e12119. https://doi.org/10.1002/eng2.12119

Patel DP, Dholakia MB, Naresh N, Srivastava PK (2012) Water harvesting structure positioning by using geo-visualization concept and prioritization of mini-watersheds through morphometric analysis in the Lower Tapi Basin. J Indian Soc Remote Sens 40:299–312. https://doi.org/10.1007/s12524-011-0147-6

Robnik-Šikonja M, Kononenko I, Marko (1997) An adaptation of Relief for attribute estimation in regression. Mach Learn Proc Fourteenth Int Conf 5:296–304

Samal DR, Gedam SS, Nagarajan R (2015) GIS based drainage morphometry and its influence on hydrology in parts of Western Ghats region, Maharashtra, India. Geocarto Int 30:755–778. https://doi.org/10.1080/10106049.2014.978903

Schumm SA (1956) Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. Geol Soc Am Bull 67:597–646

Sengupta S, Mohinuddin S, Arif M (2021) Sub-watershed prioritization for soil erosion potentiality estimation in tenughat catchment,

India. Geocarto Int 1–30. https://doi.org/10.1080/10106049.2021.2017008

Shah M, Vakharia V, Chaudhari R et al (2022) Tool wear prediction in face milling of stainless steel using singular generative adversarial network and LSTM deep learning models. Int J Adv Manuf Technol 121:723–736. https://doi.org/10.1007/s00170-022-09356-0

Singh S, Kumar A, Kumar N (2014) Motor current signature analysis for bearing fault detection in mechanical systems. Procedia Mater Sci 6:171–177. https://doi.org/10.1016/j.mspro.2014.07.021

Sonebi M, Cevik A, Grünewald S, Walraven J (2016) Modelling the fresh properties of self-compacting concrete using support vector machine approach. Constr Build Mater 106:55–64. https://doi.org/10.1016/J.CONBUILDMAT.2015.12.035

Strahler AN (1964) Quantitative geomorphology of drainage basin and channel networks. Handbook of applied hydrology

Strahler A (1997) Quantitative geomorphology. Geomorphol. Springer Berlin Heidelberg, Berlin, Heidelb

Thakkar AK, Dhiman SD (2007) Morphometric analysis and prioritization of miniwatersheds in Mohr watershed, Gujarat using remote sensing and GIS techniques. J Indian Soc Remote Sens 35:313–321. https://doi.org/10.1007/BF02990787

Upadhyay R, Padhy PK, Kankar PK (2016) Application of S-transform for automated detection of vigilance level using EEG signals. J Biol Syst 24:1–27. https://doi.org/10.1142/S0218339016500017

Vakharia V, Castelli IE, Bhavsar K, Solanki A (2022) Bandgap prediction of metal halide perovskites using regression machine learning models. Phys Lett A 422:127800. https://doi.org/10.1016/J.PHYSLETA.2021.127800

Vakharia V, Gupta VK, Kumar KP (2015a) Application of chi square feature ranking technique and random forest classifier for fault classification of bearing faults. In Proceedings of the 22th International Congress on Sound and Vibration, Florence, Italy, p 12–16

Vakharia V, Gupta VK, Kankar PK (2015b) Ball bearing fault diagnosis using supervised and unsupervised machine learning methods. Int J Acoust Vib 20 (4):244–250. https://doi.org/10.20855/ijav.2015.20.4387

Vapnik VN (1995) The nature of statistical learning theory. Nat Stat Learn Theory. https://doi.org/10.1007/978-1-4757-2440-0

Vulević T, Dragović N (2017) Multi-criteria decision analysis for sub-watersheds ranking via the PROMETHEE method. Int Soil Water Conserv Res 5:50–55. https://doi.org/10.1016/J.ISWCR.2017.01.003

Whipple KX, Tucker GE (1999) Dynamics of the stream-power river incision model: Implications for height limits of mountain ranges, landscape response timescales, and research needs. J Geophys Res Solid Earth 104:17661–17674. https://doi.org/10.1029/1999JB900120

Yadav SK, Dubey A, Szilard S, Singh SK (2018) Prioritisation of sub-watersheds based on earth observation data of agricultural dominated northern river basin of India. Geocarto Int 33:339–356. https://doi.org/10.1080/10106049.2016.1265592

Supplementaryinformation

Semiautomated tool and manual

https://drive.google.com/file/d/157VA65tjU_UdPglafaRToKh3HaoAM2RA/view?usp=sharing