



Multifactor prediction of the central Himalayan spring high-flows using machine learning classifiers

Sandipan Mukherjee³ · Sumit Sen² · Kireet Kumar¹

Received: 19 May 2022 / Accepted: 14 January 2023 / Published online: 30 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Year-round flow consistencies of the central Himalayan springs are extremely important for addressing rural water demand. As the prediction of Himalayan spring high-flows is expected to provide better opportunities for the management of excess runoff, this study aims to develop a data-driven model for predicting joint-fracture and depression type spring high-flows of the Kosi watershed of central Himalaya, India. Five machine learning algorithms are used with combinations of predictors, such as standardized anomaly of rainfall, pH, electrical conductivity and water quality index of spring water. The discharge and predictor parameters are used from a total of 06 springs distributed across the watershed, and monitored during January, 2019 to December, 2020 at monthly interval. Due to asymmetric relationships between model predictors and spring discharge, model performances are tested for the predictor time lags of 0–2 (= 60 days). A total of ten experiments are carried out, and model performances during training and testing are evaluated using receiver operator characteristics. The discriminant analysis classifier, in combination with rainfall and electrical conductivity as predictors, is found to be the best model for predicting spring high-flows.

Keywords Himalayan springs · High-flow · Prediction model · Machine learning

Introduction

The Indian Himalayan region (IHR) is a source to more than 03 million natural fresh water springs that supports approximately 50 million rural and urban population for their daily household freshwater demand (NITI Aayog 2018; Kumar et al. 2019; Panwar 2020). Springs of the IHR occur in the intersects of sloping surfaces and impermeable strata, and most of the spring discharges are through unconfined-perched aquifers with complex conduit networks that primarily depend on precipitation and recharge area

characteristics, such as subsurface geology, soil properties and dominant land-use and land-cover (Negi and Joshi 1996, 2004). Springs of the IHR are primarily joint/fracture (JT), depression (DP) and contact (CN) types, and during the last couple of decades, diminishing spring discharges in the IHR is becoming a major concern for sustainable water management (Valdiya and Bartarya 1989; Tambe et al. 2012; Daniel et al. 2021) In fact, NITI Aayog of the Government of India has surmised that approximately 50% of all the springs of IHR might have become ephemeral (NITI Aayog 2018) primarily due to climate and socioeconomical changes (Agarwal et al. 2012). As a result, acute water scarcity is becoming eminent within the region, and there is a need for field-based spring recharge activities.

To mitigate the fresh water scarcity in the IHR, sustainable water management through spring recharge activities is gaining interest among the local and governmental agencies (Rathod et al. 2021). However, the spring revival activities are often implemented with limited knowledge on the aspects of springshed geohydrology, groundwater flow regime, and linkages to water input through rainfall, resulting in limited success (Dass et al. 2021). Since spring discharges are highly variable with respect to changes in

✉ Sandipan Mukherjee
sandipan@gbpihed.nic.in

Sumit Sen
sumit.sen@hy.iitr.ac.in

¹ Centre for Land and Water Resource Management, G. B. Pant National Institute of Himalayan Environment, Kosi-Katarmal, Almora 263643, Uttarakhand, India

² Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India

³ Ladakh Regional Centre, G. B. Pant National Institute of Himalayan Environment, Leh, India

the geology, geomorphology, and rainfall climatology, irrespective of anthropogenic pressure, replication of a uniform spring revival model across the IHR is presumed to be non-productive. One of the major concerns with spring revival activities over Himalaya is the absence of sufficiently long spring discharge data. Therefore, long-term changes in the spring discharge behavior, as a response to changes in the environmental parameters, are relatively poorly quantified. The relative absence of long-term discharge and water chemistry data from different spring types has further impaired the development of numerical model/s for predicting changes in the spring flow behavior.

During the last few decades, advance nonlinear and non-stationary data analytical tools, particularly deep learning and machine learning methods are increasingly used for predicting spring discharges across the world. For example, a least-squares base lumped-parameter model was used by Zhang et al. (1996) for predicting Karst spring discharge in the USA; Granata et al. (2018) used three machine learning algorithms (i.e., regression tree, random forest, and support vector regression) for predicting Karst spring discharges of Italy. Similarly, artificial neural networks were also used by Lambrakis et al. (2000), Hu et al. (2008), Fiorillo and Doglioni (2010), and many other for predicting Karst spring discharges across the world. Recently Cheng et al. (2020) have used multilayer perceptron and memory–recurrent neural network with support vector regression for predicting spring discharges of China; and Wunsch et al. (2021) have used the 2D convolution neural networks (CNN) for extracting meteorological parameters for predicting Karst spring discharges using 1D CNN. However, due to lack of sufficiently long spring discharge and associated environmental data at a monthly or daily scale, no attempt is made until the date to develop spring discharge model for Himalaya. Since successful prediction of the spring discharge characteristics using simple geohydrological parameters is presumed to be highly beneficial for framing targeted spring revival activities for long-term sustainable water management in the IHR, this study is primarily focused on predicting JT and DP type monthly spring high-flows of a central Himalayan watershed using five very powerful and commonly used machine learning algorithms, namely Naïve Bayes classifier (NBC), k th nearest neighbor classifier (KNN), support vector machine classifier (SVM), decision tree classifier (DTC) and discriminant analysis classifier (DAC). Since year-round flow consistencies of the central Himalayan springs are extremely important for addressing rural water demand, this study is primarily focused on predicting spring-high flows, where high-flows are defined following a discharge-based criteria. Moreover, spring high-flows of the Himalayas are mostly wasted as surface run-off, not being stored or used by the local communities for use in the lean period. Subsequently, we anticipate that a proper prediction of high-flows will

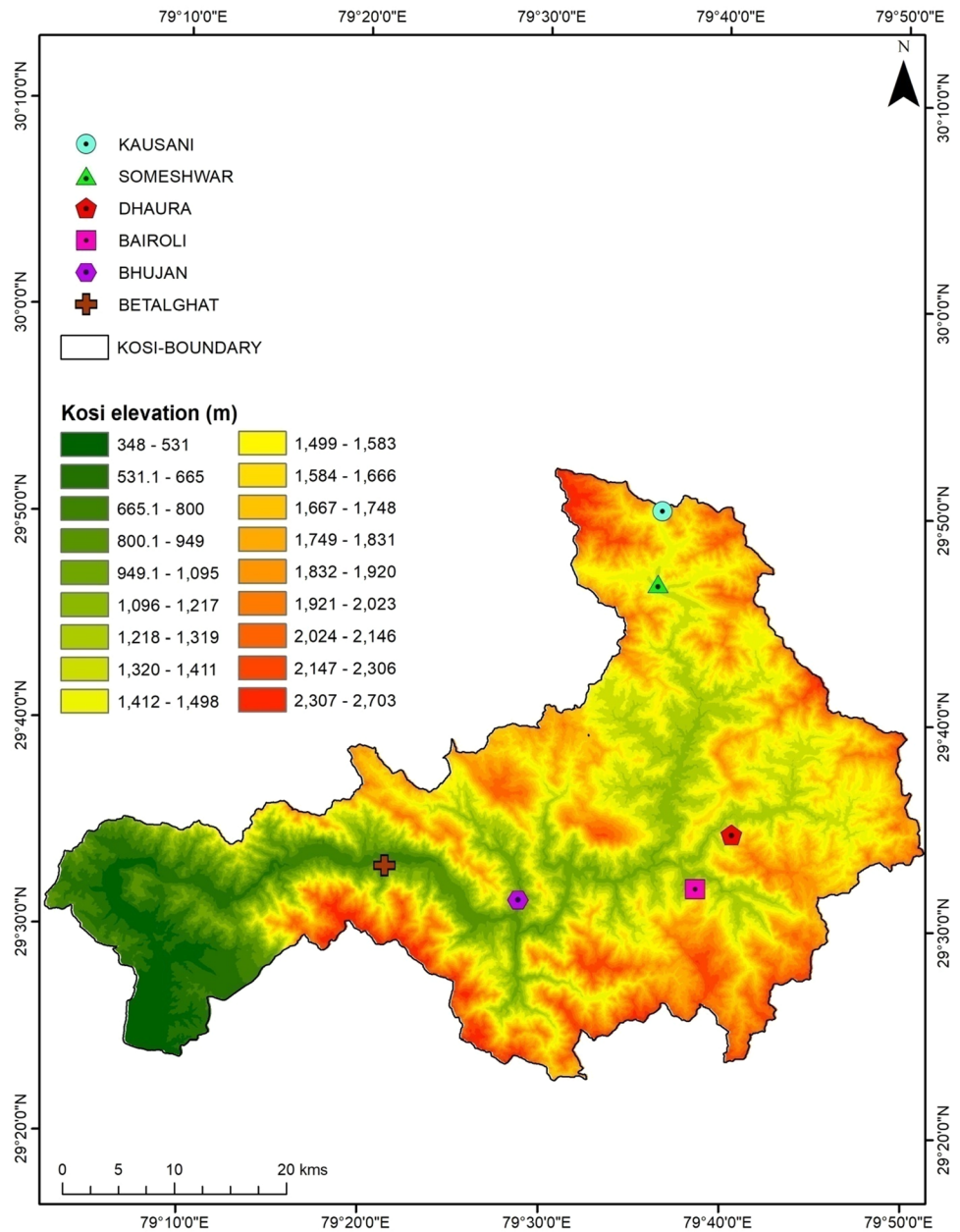
be beneficial for local communities to devise water augmentation structures for storing the excess flow for future use. Therefore, with the machine learning classifiers and a combination of simple geohydrological parameters as predictors, this study attempts to develop a data-driven intelligence model that can predict JT-DP type monthly spring high-flows of a central Himalayan watershed. The geohydrological parameters considered in this study are standardized anomalies of rainfall, pH, electrical conductivity (EC) and water quality index (WQI). The rationale for selecting these parameters for predicting spring discharge was twofold, (i) pH, EC, and WQI were assumed to be the best proxies of soil and geological factors, and (ii) rainfall was assumed to be the best control of aquifer properties and spring high flow.

Methods

Description of study area and springs

The study was carried out in the Kosi watershed of Uttarakhand, India (Fig. 1). The watershed covers an area of approximately 1863 km² wherein the major channel is the Kosi river which is stream and spring-fed with the average annual discharge varying between 0.8 and 790.0 m³ s⁻¹. Almora and Nainital are the two major districts within the watershed where an average of 50% of the total potable water supplies is made using springs, streams, and channel sources (Kumar et al. 2019). The fresh water vulnerability of Kosi watershed is high. The Takula and Ramgarh administrative blocks within the watershed are facing highest potable water scarcity (Kumar et al. 2019). The average annual rainfall of the basin is around 1200 mm, and the summer monsoon contributes around 740 mm rainfall with average numbers of heavy rainy days is 2.48 per season (Mukherjee et al. 2015, 2016; Mukherjee 2021). The region is also expected to receive higher summer monsoon rainfall and lower winter-season wet days by the end of this century under a warmer climate, hence, the natural springs and associated aquifers are expected to be affected significantly (Mukherjee et al. 2019; Ballav et al. 2021). A total of six springs (i.e., Kunagarh Dhara at Betalghat, Astola Naula at Bhujan, Bairoli Naula at Bairoli, Shiv Naula at Someswar, Jakh Naula at Dhaura, Laxmi Ashram Naula at Kausani) across the watershed were monitored, and locations of the springs, obtained through GPS (Garmin etrax30x, USA), are provided in Table 1. Among all the springs, only the Betalghat spring has surface out-flow, therefore, locally termed as ‘Dhara’. Elevations of all the springs varied from 809 to 1750 m.a.s.l. All the springs are owned by local communities, and spring water is used for daily household activities. The dominant land-use land cover of each spring is provided in Fig. 2 for an area circumscribing 500 m upslope. It is

Fig. 1 Location of springs over the Kosi watershed



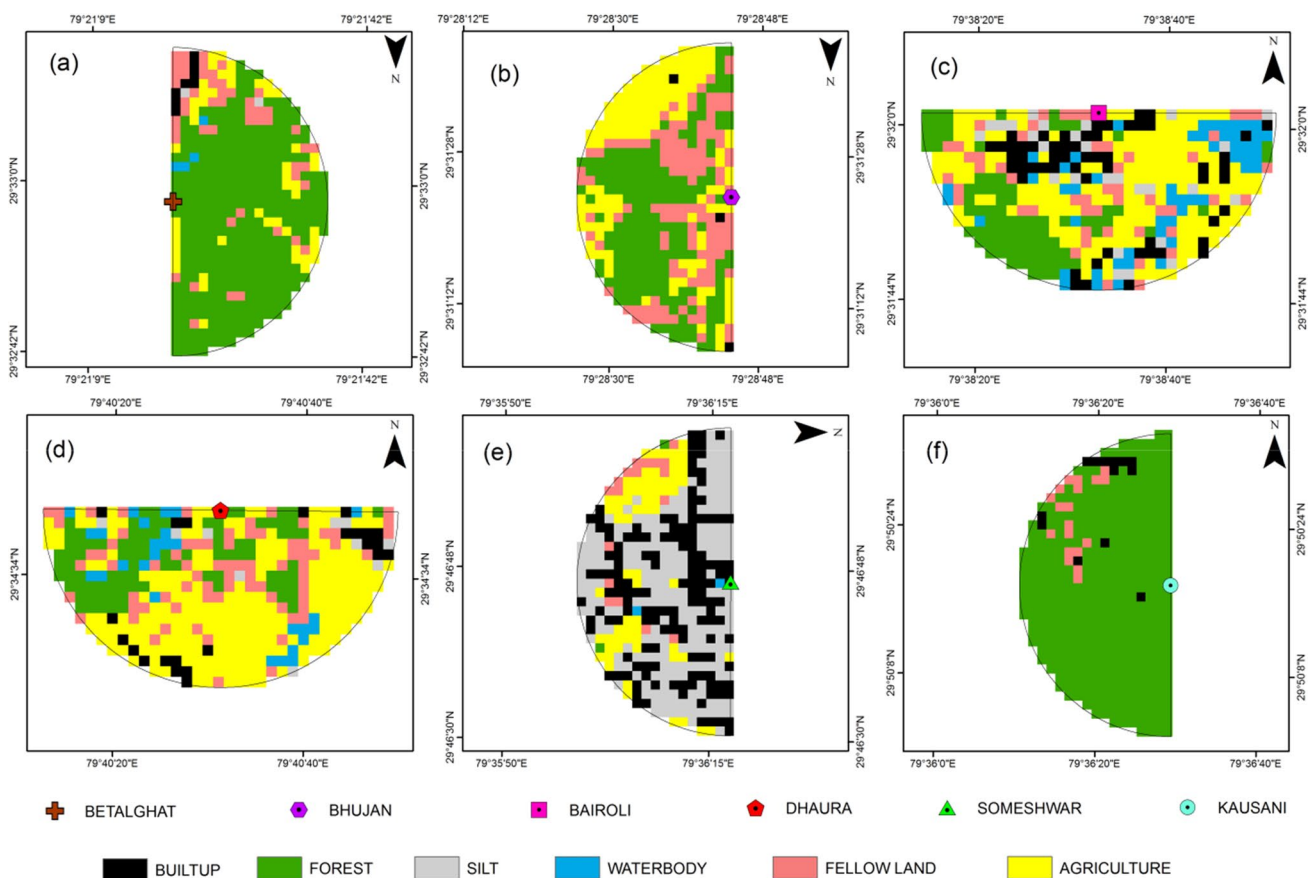
noted from Fig. 2 that except for Someswar spring, which is dominated by silt, the rests of the springs were mostly dominated by agricultural lands or forested vegetation. Similarly, except for the Someswar and Kausani springs, which are depression (DP) springs, the rest of the springs are joint/fracture (JT/FR) in nature. Springs at the higher elevations of the Kosi watershed, i.e., at Dhaura and Kausani, are dominated by quartzite and mica schist, whereas the Betalghat spring at 803 m is dominated by dolostone ($\text{CaMg}(\text{CO}_3)_2$).

Data description

All six springs across the watershed were monitored at a monthly interval during January, 2019 to December, 2020. Monthly discharges of these springs were computed by averaging single-day observations carried out at 1000, 1300, and 1600 h Indian Standard Time (IST). The discharges were manually measured using stopwatch and a measuring cylinder as indicated by Negi and Joshi (2004). Two sample *t* tests were carried out using discharge data between the pairs of all the springs to identify variability of mean flow. The null hypothesis considered for the *t*-test was that the pair-wise difference between discharges of two springs has a mean

Table 1 Descriptions of springs used for monthly data collection

Name	Location	Elevation (m)	Aspect [slope range]	Spring type	Dominated rock type	Q_{60} L m ⁻¹
Kunagarh Dhara at Betalghat	29° 32' 58.0" N 79° 21' 19.1" E	809	Westerly [21.4°–23.7°]	FR/JT	Dolostone	4.03
Astola Naula at Bhujan	29° 31' 23.6" N 79° 28' 44.6" E	931	South-easterly [14.4°–16.7°]	FR/JT	Quartzite	0.18
Bairoli Naula at Bairoli	29° 32' 01.3" N 79° 38' 32.8" E	1331	Southerly [14.4°–16.7°]	FR/JT	Gneiss	17.23
Shiv Naula at Someswar	29° 46' 46.7" N 79° 36' 17.6" E	1381	South-easterly [5°–7.3°]	DP	River deposit upon Quartzite	14.13
Jakh Naula at Dhaura	29° 34' 40.1" N 79° 40' 31.2" E	1530	Southerly [5°–7.3°]	FR/JT	Mica schist and Quartzite	0.64
Laxmi Ashram Naula at Kausani	29° 50' 17.8" N 79° 36' 29.4" E	1749	South-westerly [12°–14.4°]	DP	Quartzite	1.02

**Fig. 2** Land-use land cover maps of each selected spring for an area circumfusing 500 m upslope. Subplots (a–f) are for Betalghat, Bhujan, Bairoli, Dhaura, Someswar, and Kausani, respectively

equal to zero at 95% confidence interval. A rejection of the null hypothesis would indicate that the discharges of both springs are having nonzero mean and are governed by asymmetric processes. We found that among the 15 combination pairs of discharges, null hypothesis was always rejected at 95% confidence interval indicating that the discharge

characteristic of each spring was unique. Electrical conductivity (EC, $\mu\text{S cm}^{-1}$) and pH of each sample was measured instantaneously using pHep Tester (accuracy = ± 0.1 pH, Hanna Instrument, Romania) and DiST3-HI98303 EC Tester (accuracy = $\pm 2\%$ full scale, Hanna Instrument, Romania).

Subsequently, monthly averages of discharge, pH and EC were produced.

In order to carry out detail physico-chemical analysis of spring-water, similar to discharge measurements, samples were collected in sealed 500 ml containers at 1000, 1300, and 1600 h IST at the day of sampling, and analyzed in the Central Laboratory Facility of GBPNIHE, Almora, India, for sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), chloride (Cl^-), sulfate (SO_4^{2-}), alkalinity as HCO_3^- , and nitrate (NO_3^-) in mg L^{-1} . The instrument details were similar to Rani et al. (2021) and standard laboratory practices as per Bureau of Indian Standard (BIS 10500 2012) were followed. Subsequently, monthly averages were produced. Data could not be collected for a total of three months due to logistical issues. Discharge, pH, EC, and water chemistry data for these months were gap-filled using simple linear interpolations. Since rain gauges were not operational at each spring location, rainfall data near to each spring location was extracted from the monthly gridded rainfall products of CHIRPS version 2.0 of Climate Hazards Group Infra-Red Precipitation with Stations having a spatial resolution of 5×5 km (<https://www.chc.ucsb.edu/data/chirps>). There were a total of 06 pixels of rainfalls near to 06 springs were used. The rationale for using CHIRPS data was its suitable conformation to observation over Uttarakhand state of India (Banerjee et al. 2020).

Hydrological analyses

To comprehend the basic hydrological characteristics of each spring considered in this study, auto-correlation (*acf*) and cross-correlation (*ccf*) analyses with rainfall were carried out followed by assessment of linkages between spring water chemistry and rainfall. Finally, flow duration curve (FDC) was produced for each spring to assess the cumulative response of geo-hydrological factors affecting flow regime and to identify the high-flow threshold. As indicated by Mayaud et al. (2014) and Dass et al. (2021), the *acf* provides an estimate of memory of the system whereas the *ccf* would indicate linear relationships between discharge and rainfall. Therefore, *acf* and *ccf* information was used to extract information on aquifer behavior. The *acf* for each spring discharge time series was computed for time lag 1–12 to consider the annual signatures. Similarly, the *ccf* for each spring was estimated using discharge and rainfall time series for a time lag period of 1–12 months. The time lag information where statistically significant *ccfs* were obtained were further used in the high-flow prediction models. Once the uni- and bivariate analyses were carried out, FDC of individual spring was produced to assess the flow characteristics of each spring irrespective of occurrences. Subsequently, frequency distribution plots of flows of all the springs were produced and high-flow events were identified using

threshold discharge values. Identified high-flow threshold of each spring was further used for dichotomous categorization of flow time series used for prediction model. Water chemical properties of all the springs were carried by categorizing chemical facies using piper diagrams. The monthly water quality indices (WQI) were produced for each spring using the methods of Brown et al. (1972) and Gebrehiwot et al. (2011) where the WHO standards for drinking water of Na^+ , K^+ , Ca^{2+} , Mg^{2+} , Cl^- , NO_3^- , SO_4^{2-} , and pH were considered. The WQI values were further considered for predicting spring high-flows.

Classifiers and applications

The multifactor prediction of spring discharge was carried out using five very powerful commonly used machine learning classifiers, i.e., NBC, KNN, SVM, DTC and DAC. Brief descriptions of all these models with application procedures are provided below.

Classifiers

The NBC is a simple but powerful machine learning algorithm developed on the principal of Bayesian probabilistic classifications wherein posterior probabilities of a hypothesis are obtained by combining data based on new evidences with prior probabilities of the same hypothesis. The NBC constructs a Bayesian model assigning probability of posterior class to an instance as: $P(Y = y_j | X = x_i)$, where it could be assumed that a data are composed of n cases as x_i , with $i = 1$ to n having p attributes as $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. Further, $y \in \{y_1, y_2, \dots, y_c\}$ indicating each instance belongs to only one class. As indicated in Berrar (2019), machine learning predictive models produce a score (s) for each instance of x_i which is used to quantify the degree of class memberships of a certain case in y_j . In case of positive or negative instances such that $y \in \{0, 1\}$, a predictive model can be classified as a ‘ranker’ or a ‘classifier’. In a ranker, scores are used to order instances from most likely to most unlikely. The ranker can be considered as a classifier when a threshold (t) is used in a ranking score such that $\{s(x) \geq t\} = 1$. An instance is assigned to a class in a NBC using $P(Y = y_j | X = x_i)$. Therefore, it is assumed in NBC that predictors of a model are conditionally independent when class information is available. Subsequently, observations are assigned to most likely classes by the NBC, and density of predictors within each class is estimated with model posterior probabilities. The conditional probabilities of NBC classifier were estimated in this study using the kernel density estimation as the kernel density estimation demonstrates very good efficacy for imbalanced data (Murakami and Mizuguchi 2010).

The KNN classifier is a supervised machine learning non-parametric algorithm which is initiated through searching a

specific number of nearest neighbor within the training data set. The KNN classifier works on the principle of calculating distances on n number of characteristics. The distance calculation could be Euclidean, Manhattan or other. Therefore, unlabeled observations are assigned to a class which is most similar to a labeled class. The model performance is highly dependent on the choice of number of neighbors, and the value of number of neighbors should be such that the model does not over or under fit data (Zhang 2016). The number of neighbor for fitting the training data of this study was kept to be 17 after a small sensitivity study associated to best performance of the model.

Similar to the KNN, SVM is a supervised machine learning algorithm, and data are classified by the SVM by finding the best hyperplane separating two different classes (Gutierrez 2015). Therefore, a best hyperplane would have highest difference with no data points between two classes. The support vectors are those data points that are closest to the class separating hyperplane. The SVM are highly versatile and memory efficient. The SVM can have the Gaussian radial basis function (RBF), linear kernel and user customized kernels. However, the kernel function used in this study was RBF as it is found to perform better than the linear kernel (Savas and Dervis 2019).

The DTC is also a nonparametric supervised machine learning algorithm. The DTC predicts the target variable using simplified decision rules of training data. The root is considered as the base of the tree, subsequently, multiple branches are designated as nodes. The categorical separations of classes are carried out at each node (Namous et al. 2021). Algorithm for finding best split in the categorical predictor was used in this study using $2^{c-1} - 1$ combinations of c categories.

The linear discriminant analysis classifier (DAC) is widely used in categorization studies. The DAC classification technique is applied with the assumption that data distributions of different classes are Gaussian. Therefore, the classifier function estimates parameters of Gaussian distribution of each class during training. Class categorization using DAC is being carried out by searching for smallest misclassification cost function, and the method assumes common population covariance matrix for all the classes (Sifaou et al. 2020).

Application procedures

To implement the classifiers, the high-flows of all the springs were identified from their FDCs (details of FDCs are provided in “Flow duration curves (FDC)”). The Q_{60} discharge value of each FDC was considered as the threshold or lower limit of spring high-flow implying that any discharge value greater than Q_{60} value of each spring was considered as high-flow. The Q_{60} threshold limit was 10% higher than the

50% sustainable flow limit of Tarafdar (2013) identified for the springs of central Himalaya. Once the high-flow threshold is identified for each spring (Table 1), the discharge time series for each spring was converted into a dichotomous series of H and L , where H indicates high-flow and L indicates low-flow. Finally, a dichotomous series with 144 samples with H and L was produced by combining all the discharge time series. A total of 74 values were classified as L , whereas the remaining 70 values were H . This 144 sample H and L time series was used as predictand for the models.

To find out the best data driven intelligence model that can predict JT-DP type spring high-flows a combination of 10 numbers of experimental simulations were carried out using standardized anomaly of rainfall, pH, EC, and WQI as predictors (Table 2). Standardized anomaly of pH (pH_a), EC (EC_a) and rainfall (R_a) was computed for each parameter of individual spring as: $D_a = (D_i - \bar{D})/\sigma_D$, where D_i was the monthly value, \bar{D} was the monthly average value during 2019–20, and σ_D was the standard deviation of monthly value. Since WQI was inherently scaled, no standardized anomaly of the WQI was estimated. To obtain a robust model, each simulation was carried out using bootstrapping method where 2/3rd of the all the data points, i.e. 96 samples were randomly used to train the model and the rests 48 random samples were used for model validation. The bootstrapping procedure was applied for 100 times. Finally, ensemble mean of model predictions for training and testing data was computed using receiver operating characteristics (ROC) having 1 – specificity and sensitivity, where specificity included selectivity or true negative rate, and sensitivity included hit rate or true positive rate (Fawcett 2006). The rationale for using ROC analysis, rather than the traditional ANOVA test/ t test, for evaluating the model performances of this study, is that we converted our data into a

Table 2 Total number of experiments with combinations of predictors, time lag used for each predictor and classifiers used for model development are indicated

Experiment	Predictor/s	Time lag	Classifiers
E1	Rainfall anomaly (R_a)	Lag=0–2	NBC, KNN, SVM, DTC, DAC
E2	pH anomaly (pH_a)	Lag=0–2	– Do –
E3	EC anomaly (EC_a)	Lag=0–2	– Do –
E4	WQI	Lag=0–2	– Do –
E5	$R_a + pH_a$	Lag=0–2	– Do –
E6	$R_a + WQI$	Lag=0–2	– Do –
E7	$pH_a + WQI$	Lag=0–2	– Do –
E8	$R_a + pH_a + WQI$	Lag=0–2	– Do –
E9	$R_a + pH_a + EC_a$	Lag=0–2	– Do –
E10	$R_a + EC_a$	Lag=0–2	– Do –

dichotomous series (or binary classifier) of high flow and non-high flow. Consequently, the best evaluator criterion for the model performance is accuracy, which is represented in the ROC analysis through Sensitivity (true positive rate) and Specificity (true negative rate). On the contrary, ANOVA and traditional t-test is focused on assessing how well the mean of a variable is modeled, which is not the primary intention of the model verifications of this study. As the area under ROC curve (AUC) shows model accuracy predicting dichotomous events (Negnevitsky 2001), the prediction accuracy classification based on the highest AUC was used to evaluate each model performance.

Results and discussions

This section is organized such that “Spring hydrodynamics”, “Flow duration curves (FDC)” and “Seasonal water chemistry” describe individual spring geo-hydrological properties, nature of the flow duration curves and variation in the spring water chemical constituents. “The high-flow predictions” elaborates details of the numerical modeling carried out for categorizing high-flow as a response to rainfall.

Spring hydrodynamics

A spring hydrograph is expected to provide recharge and storage potential with aquifer transmissive properties through retardation response of rainfall. Furthermore, a spring hydrograph could be similar to surface stream

hydrographs if the aquifers are unconfined and have quick response time to rainfall input (Kresic and Bonacci 2010). The spring hydrographs were primarily analyzed with respect to general retardation response to rainfall throughout the observation period.

All the selected springs of this study are perennial, and their monthly discharges (Q) varied between 0.146 and 19.76 L m⁻¹ (lpm). The monthly average highest discharge was noted for the Bairoli spring (18.22 L m⁻¹), whereas the monthly average lowest discharge was noted for the Bhujan spring (0.288 L m⁻¹). Irrespective of springs, the median discharge was 1.08 L m⁻¹. The average annual discharges of springs at Betalghat, Bhujan, Bairoli, Someswar, Dhaura, and Kausani were noted to be 7.34, 0.28, 18.2, 15.4, 0.65, and 1.08 L m⁻¹, respectively. Therefore, as per the modified Meinzer classification of springs (Meinzer 1927; Springer et al. 2008), Bhujan and Dhaura springs were fifth order, Betalghat, Someswar, and Kausani springs were sixth order, and Bairoli spring was seventh order. Irrespective of spring locations, the average monsoon period rainfall during 2019–20 was 971.9 mm which was around 231.9 mm higher than the long-term average of 740 mm (Mukherjee et al. 2015). The rainfall amount had asymmetric impact on the discharge properties of springs, and the same can be noted in Fig. 3, and it can be noted that the Betalghat spring had the fastest response to rainfall. Further impacts of rainfall on discharges are analyzed using *ccfs* below.

The *acf* of individual spring discharges are provided in Fig. 4a. In order to identify the generalized signatures of *acfs*, the ensemble averages were also produced. The *acf* values of all the springs were approximately equal to 0 at

Fig. 3 Monthly average discharges (L m⁻¹) of springs are presented with monthly total rainfall (mm) during January 2019 to December 2020. The vertical lines are standard errors

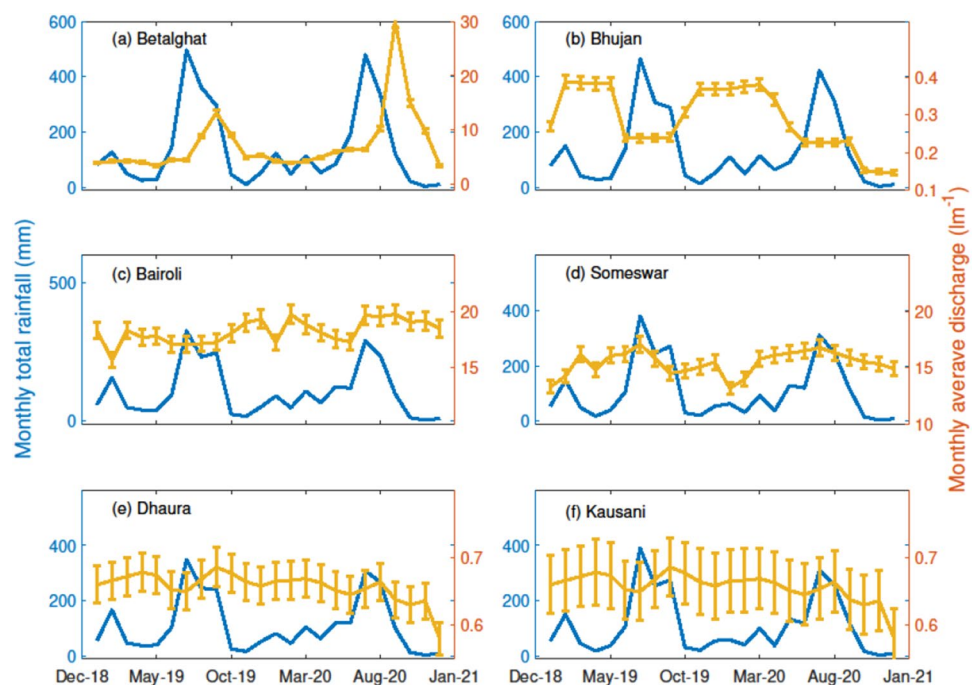
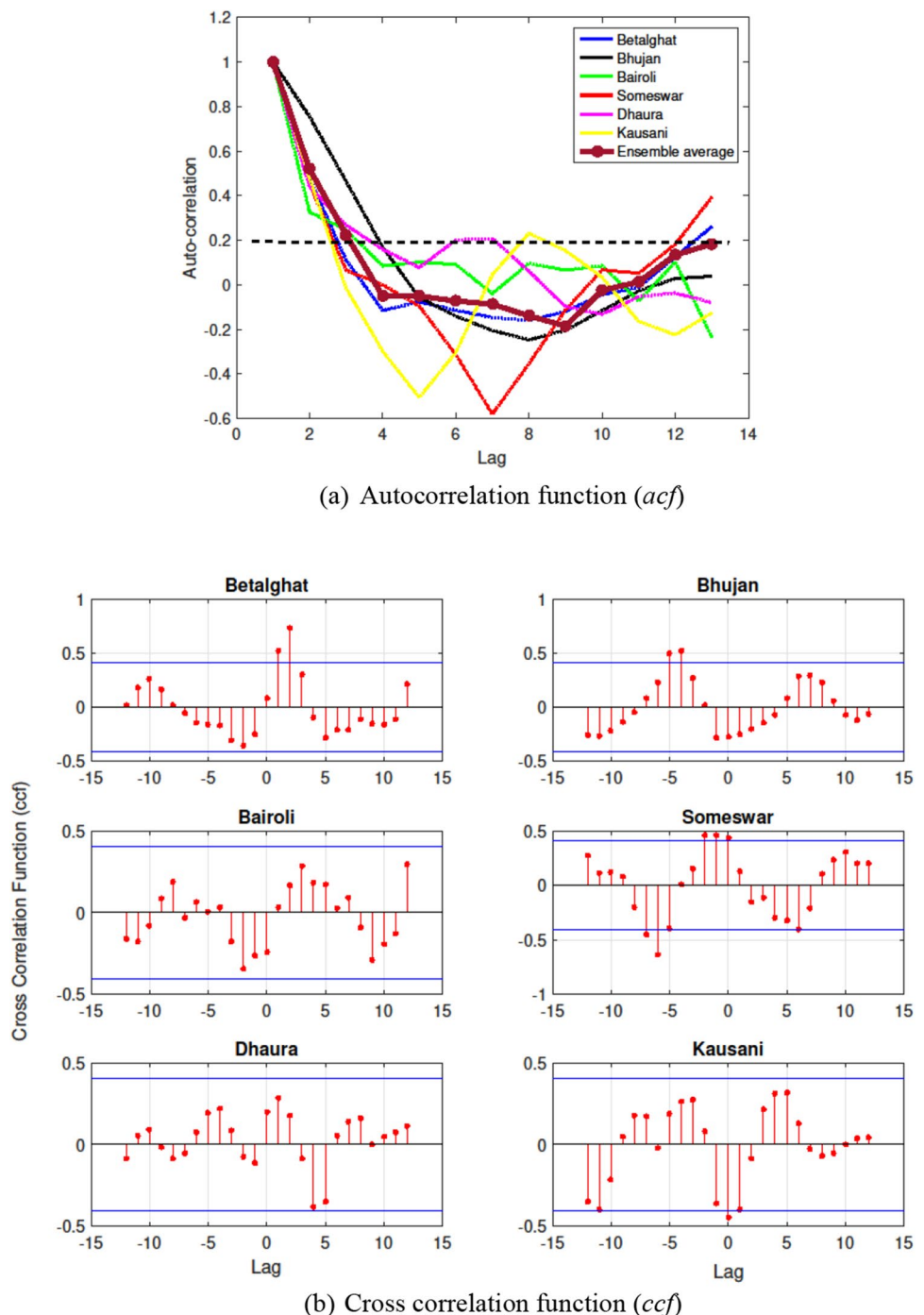


Fig. 4 Auto-correlation functions (*acf*) and cross-correlation functions (*ccf*) of each springs are provided in subplots (a) and (b). The dashed line at $acf=0.2$ indicates statistically accepted value as of Mayaud et al. (2014)



time lag 4 (= 120 days); whereas the positive ‘memory effect’, that is when $acf=0.2$ after Mayaud et al. (2014), was noted for an approximate lag of 2 (= 60 days) when all the springs were considered. The maximum $acf=0.75$ at lag 1 (= 30 days) was noted for Bhujan spring, whereas lowest $acf=0.32$ at lag 1 (= 30 days) was noted for Bairoli spring. However, if the rate of change of *acf* with time lag is considered till $acf=0$, fastest diminishing auto-correlation was observed for the Kausani spring followed by Someswar and

Betalghat springs indicating relatively lower storage capacities of aquifers contributing to these springs. On the contrary, Bairoli spring had the lowest diminishing rates of *acf* indicating relatively prolonged memory effects and higher storage capacities of aquifers. As indicated by Dass et al. (2021), fastest diminishing auto-correlation of the Kausani, Someswar and Betalghat springs could be linked to bigger fractures with broader network of flow paths and evolved geological fractures resulting relatively lesser memory

effects. On the contrary, relatively higher memory effects of Bairoli springs could be linked to smaller fractures with narrower drainage network and constrained geological fractures.

The *ccf* of individual spring discharge and total monthly rainfall is provided in Fig. 4b. It can be noted from the *ccfs* that the time delays between monthly rain and discharge vary significantly. Asymmetric relationships between rainfall and discharges at 95% CI, indicating high dependency of springshed output to rainfall input, could be noted for Betalghat and Someswar springs where rainfall resulting enhancement in the discharges within a monthly lag period of 0–2 (=0–60 days). The maximum *ccfs* of Betalghat and Someswar were 0.74 and 0.49 at 95% CI, respectively. However, relative dependency of springshed output to rainfall input was also noted for the Bairoli and Dhaura springs having highest *ccfs* of 0.29 and 0.285 at lag periods of 03 and 01 months, respectively, but not at 95% CI. Rest of the two springs, i.e. Bhujan (min *ccf* = -0.28 at lag time = 0) and Kausani (min *ccf* = -0.44 at lag time = 0), were not having any significant relationship between rainfall and discharge. Therefore, order of rainfall on discharges could be summarized as: Betalghat > Someswar > Bairoli > Dhaura. Highest dependency of the Betalghat spring to rainfall could be linked to highest terrain slopes (=21.4°–23.8°) among all the springs whereas the Kausani spring, having lowest dependency to rainfall, could be corroborated to large forest cover surrounding the spring (around 93%) with lower terrain slopes (=12°–14.4°). Since, the *ccfs* indicated monthly rainfall impacted discharge within a time lag period of 0–2 months, irrespective of springs, the predictive models were tested for 0–2 month time lag of each predictor.

Flow duration curves (FDC)

The FDCs were analyzed with respect to (i) slope characteristics where a steep slope indicates highly variable flow regime and a flatter FDC indicating aquifer storage with gradual discharge; and (ii) overall (Q_{10}/Q_{90}) and high flow (Q_{60}/Q_{90}) variability. Further, Q_{10}/Q_{90} values were used to compare flow variability ratios following (Alfaro and Wallace 1994). FDCs of all the springs are produced in Fig. 5. It can be noted from the general distribution of FDCs that flow properties of Bairoli, Someswar, and Dhaura springs were almost similar with moderately flat slopes sustained by the ground water discharges. Therefore, the Q_{mx}/Q_{mn} ratios (where, Q_{mx} , Q_{mn} represents maximum and mean discharges during the observation period) of these three springs were almost identical, 1.08, 1.10, and 1.04, respectively. Flat slopes in the upper-end of the FDCs were noted for all the springs, except for Betalghat, indicating that deep aquifers were mostly contributing to discharges, and these aquifers might evacuate, if not recharged through precipitation and ground water augmentation activities. Similarities in the

upper-end of FDCs were also indicating that the physiographic characteristics, particularly, sub-surface drainage distribution patterns of all five springs, except Betalghat, were comparable. Consequently, it can be stated that any spring revival activities should incorporate detail analyses of slope-characteristics of FDCs before ground augmentation. Except for the Betalghat spring ($Q_{10}/Q_{90}=3.4$) categorized as ‘moderately (well) balanced’, rests of all the springs were found to be ‘extraordinarily balanced’, i.e., Bhujan ($Q_{10}/Q_{90}=2.12$), Bairoli ($Q_{10}/Q_{90}=1.14$), Someswar ($Q_{10}/Q_{90}=1.18$), Dhaura ($Q_{10}/Q_{90}=1.06$), and Kausani ($Q_{10}/Q_{90}=1.15$), respectively (Meinzer 1927; Alfaro and Wallace 1994). The high flow variability of each spring was further quantified using Q_{60}/Q_{90} values. Although the Betalghat spring was noted to have highest overall flow variability, the high-flow variability was lowest for the Dhaura spring ($Q_{60}/Q_{90}=1.03$) and highest for the Bhujan spring ($Q_{60}/Q_{90}=1.31$), indicating that the poor flow consistency of the Bhujan spring. The Q_{60} threshold value of each spring was provided in Table 2 and was used for producing the dichotomous time series of *H* and *L* flow and used for multifactor flow characterization.

Seasonal water chemistry

An earlier study on the spring water chemistry of upper Kosi watershed by Rani et al. (2021) has indicated that two types of hydro-chemical complexions were dominant within the watershed during pre-monsoon summer periods, $Ca^{2+}-Mg^{2+}-HCO_3^-$ and $Ca^{2+}-Mg^{2+}-Cl^-$. The $Ca^{2+}-Mg^{2+}-Cl^-$ complexions were reported to change to $Ca^{2+}-Mg^{2+}-HCO_3^-$ during post monsoon. However, irrespective of the spring locations and seasons, the general water chemical signatures (Fig. 6) indicated weak acid (HCO_3^-) exceeded strong acids (SO_4^{2-} and Cl^-) for all the six springs. Similarly, alkaline earths (Ca^{2+} and Mg^{2+}) exceeded alkali (Na^+). Therefore, it can be inferred that the dominant complexion for the six study springs was $Ca^{2+}-Mg^{2+}-HCO_3^-$ which did not show much variation with season. One of the significant features of Fig. 6 was that the Dhaura spring had no dominant cation type. Similarly, monsoon rainfall was noted to minutely convert the $Mg^{2+}-Cl^-$ complexion to $Ca^{2+}-Cl^-$ for the Betalghat spring. On the contrary, monsoon rainfall resulted $Ca^{2+}-Cl^-$ complexion of Bhujan spring to be partially converted to be $Mg^{2+}-Cl^-$. Although rock outcrop of the Betalghat and Bhujan springs clearly indicated the presence of dolostone and quartzite, the subsurface geology of Bhujan spring clearly indicated presence of dolomite with $Ca/Mg=1.36$ (Table 3), which is in corroboration to Ca/Mg ratio near to unity for dolomite dominated springs (White 2010). Consequently, the average annual EC of Bhujan spring was noted to be highest (451.6 $\mu S\ cm^{-1}$) among all the springs. As the average

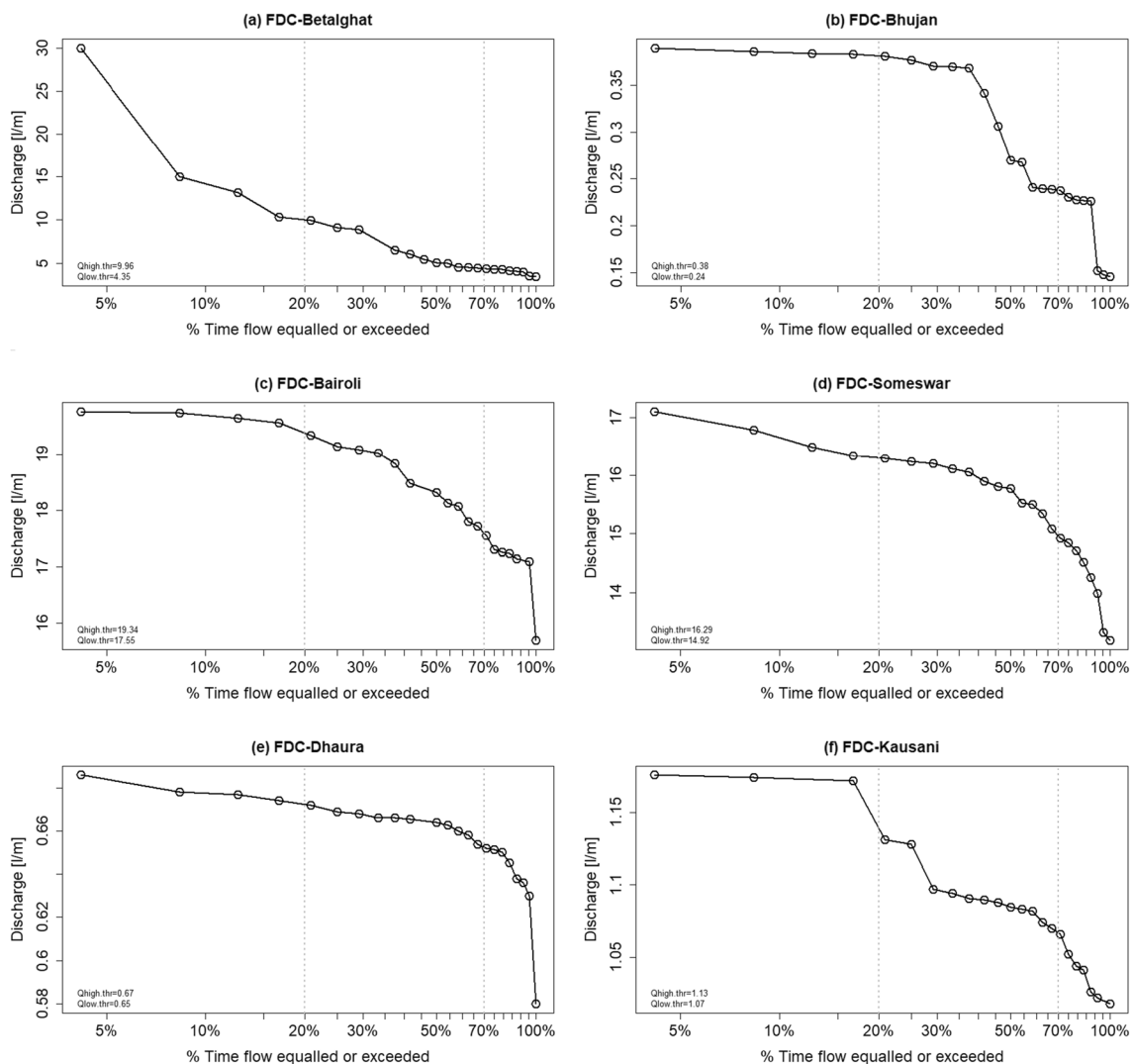


Fig. 5 Flow duration curves (FDC) of all the study springs are produced using monthly discharge data

annual NO_3^- concentration was comparatively higher for the Someswar spring (6.99 mg L^{-1}), it can be inferred that the anthropogenic contamination through leaching of sewage water might be the reason. However, the annual average pH levels of all the study springs (6.79–7.57) were within the potable range as per BIS. The average WQI values of the Betalghat, Bhujan, Bairoli, Someswar, Dhaura, and Kausani springs were $25.0 (\pm 2.9)$, $38.2 (\pm 5.7)$, $17.3 (\pm 1.8)$, $31.9 (\pm 5.0)$, $17.4 (\pm 2.7)$, and $19.4 (\pm 2.6)$, respectively, where values in parentheses are standard deviations. Since, WQI values of all the springs were below 50, generic water quality of all the springs could be categorized as ‘excellent’ as per the BIS standards of India. The cumulative mean WQI time series with 144 data were further used for multifactor flow predictions.

The high-flow predictions

The spring high-flow predictions were carried out using five machine learning classifiers for a total of ten experiments (E1–10) having different combinations of predictors with predictor time lags of 0–2 months. As indicated in “[Application procedures](#)”, a total of 96 samples were randomly used to train the model and 48 random samples were used for model testing. The bootstrapping simulations were carried out for 100 times. The model performances during training and testing of each experiment could be found in Figs. 7 and 8, respectively. In view of the extreme variability of JT-DP type spring flow properties, any AUC value between 0.6 and 0.7 was considered as ‘acceptable’ and AUC value greater than 0.7 was considered as ‘good’ (Moriassi et al. 2007). Firstly, the individual experiment results were analyzed by highlighting the best

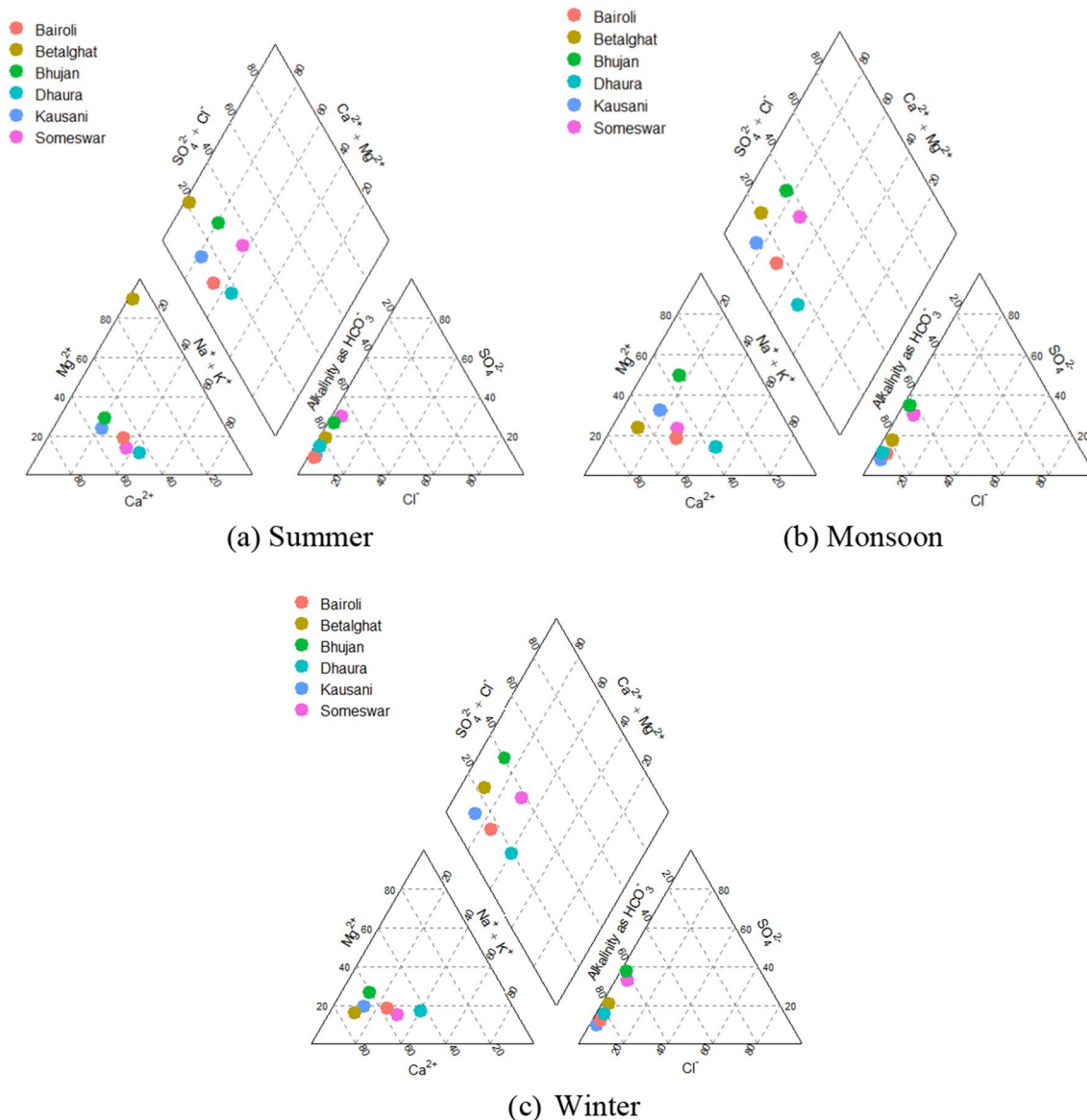


Fig. 6 Piper plots of spring water ions during **a** summer (March–May), **b** monsoon (June–September) and **c** winter seasons (November–February)

predictor/s and best time lag based on mean AUC values. Finally, the best classifier for predicting the JT–DP spring high-flow over Kosi-watershed was selected by identifying highest mean AUC during testing of an experiment at a certain time lag of predictors.

During the training and test simulations of the NBC, highest mean AUCs were noted to be 0.68 (± 0.05) and 0.68 (± 0.07), respectively, when WQI was the predictor (experiment E4) at lag = 2. Similarly, mean AUCs for training and testing of KNN were 0.66 (± 0.02) and 0.67 (± 0.05), respectively, when $R_a + pH_a + EC_a$ were predictors (experiment E9) at lag = 2. Mean AUCs for training and testing of SVM were 0.58 (± 0.01) when EC_a was predictor (experiment E3)

at lag = 1, and 0.62 (± 0.05) when $R_a + EC_a$ was predictor (experiment E10) at lag = 2. Therefore, identification of a best combination of predictors at a specific time lag using the SVM was inconclusive. However, training and testing results of the DTC was conclusive, and mean AUCs for training and testing were 0.67 (± 0.03) and 0.67 (± 0.05), respectively, when EC_a was predictor (experiment E3) at lag = 2. Performance of the DAC was best among all the classifiers during training and test. Moreover, the impacts of changing predictors and time lag on the DAC performances were not significant as the average AUCs for training and testing, irrespective of predictors and time lag, were 0.70 (± 0.002) and 0.71 (± 0.003), respectively. However, a finer

Table 3 Annual mean values of the physicochemical parameters of the spring water samples of six sites

Site	pH	EC ($\mu\text{S cm}^{-1}$)	Sodium (mg L^{-1})	Potassium (mg L^{-1})	Calcium (mg L^{-1})	Magnesium (mg L^{-1})	Chloride (mg L^{-1})	Nitrate (mg L^{-1})	Sulfate (mg L^{-1})	Alkalinity (mg L^{-1})
Betalghat	7.55	108.64	4.06	0.86	25.88	7.69	2.63	1.84	14.84	59.86
Bhujan	7.57	451.63	10.23	1.84	37.04	27.10	2.62	1.98	31.51	58.88
Bairoli	6.99	74.68	6.41	1.06	13.74	4.98	2.10	1.05	5.81	47.74
Someshwar	6.79	313.91	12.46	10.10	36.45	12.73	3.94	6.99	21.52	43.79
Dhaura	7.30	54.24	6.75	1.05	7.29	2.55	1.83	0.39	8.15	49.57
Kausani	7.06	124.98	6.13	1.18	25.50	11.01	2.21	0.39	5.57	52.95

inspection of the DAC performance revealed best performance was achieved when $R_a + EC_a$ was used as predictors (experiment E10) at time lag = 0, i.e., mean AUC during testing was 0.72 (± 0.02). Hence, except for the SVM, ‘acceptable’ performances were noted for NBC, KNN, and DTC, and ‘good’ performance was noted for DAC. Finally, it can be concluded that JT-DP type monthly spring high-flows could be predicted with acceptable confidence using standardized anomaly of rainfall and electrical conductivity with no time lag using the discriminant analysis classifier.

Summary and conclusions

A spring discharge prediction model is an extremely important tool for assessing long-term flow sustainability. The nonlinear and non-stationary data analytical methods provide excellent opportunities for developing such models wherein simple geo-hydrological parameters are used as predictors. However, in spite of life-supporting roles of the Indian Himalayan springs, all most no efforts are made to develop a spring discharge prediction model for the Indian Himalayas. In order to address this knowledge gap, this current study was aimed at developing a simple machine-learning-based model that could predict JT and DP type monthly spring high-flows of a central Himalayan watershed. Subsequently, five very powerful machine learning algorithms (NBC, KNN, SVM, DTC, and DAC) were used with combinations of simple predictors, such as standardized anomaly of rainfall, pH, EC and WQI. Spring discharges, pHs, ECs and spring water chemical properties were monitored during 2019–2020 on monthly basis from 06 numbers of JT-DP spring (a cumulative of 144 data points for each parameter) of Kosi-watershed, Uttarakhand, India. All this six springs considered in this study are perennial and owned by the local community for daily water usages. Monthly discharges of these springs varied between 0.146 and 19.76 L m^{-1} , and flow properties were either ‘moderately (well) balanced’ or ‘extremely well balanced’ as per the modified Meinzer classification. The annual average pH levels of all the springs (6.79–7.57) were within the potable range, and the dominant spring water complexation was $\text{Ca}^{2+}\text{-Mg}^{2+}\text{-HCO}_3^-$ having little variation with season. Asymmetric relationships between rainfall and discharges at 95% CI were noted when all the springs were considered, and rainfall resulting enhancement in the discharges up to a monthly lag period of 2 (= 60 days) were noted in few cases. As a consequence, model performances were tested for time lags 0–2. The spring high-flows were identified using the Q_{60} value of the FDC of each spring that varied between 0.18 and 17.23 L m^{-1} . Except for the WQI, standardized anomaly of rainfall, pH, and EC were used as predictors. A total of ten

Fig. 7 The mean area under curve (AUC) with standard deviations obtained using the receiver operating characteristics (ROC) of all the training experiments for predictor(s) lag = 0–2 months are presented. The mean and standard deviation values are obtained using 100 bootstrapped samples. *NBC* Naïve Bayes, *KNN* *k*th nearest neighbors, *SVM* support vector machine, *DTC* classifier, *DAC* discriminant analysis classifier

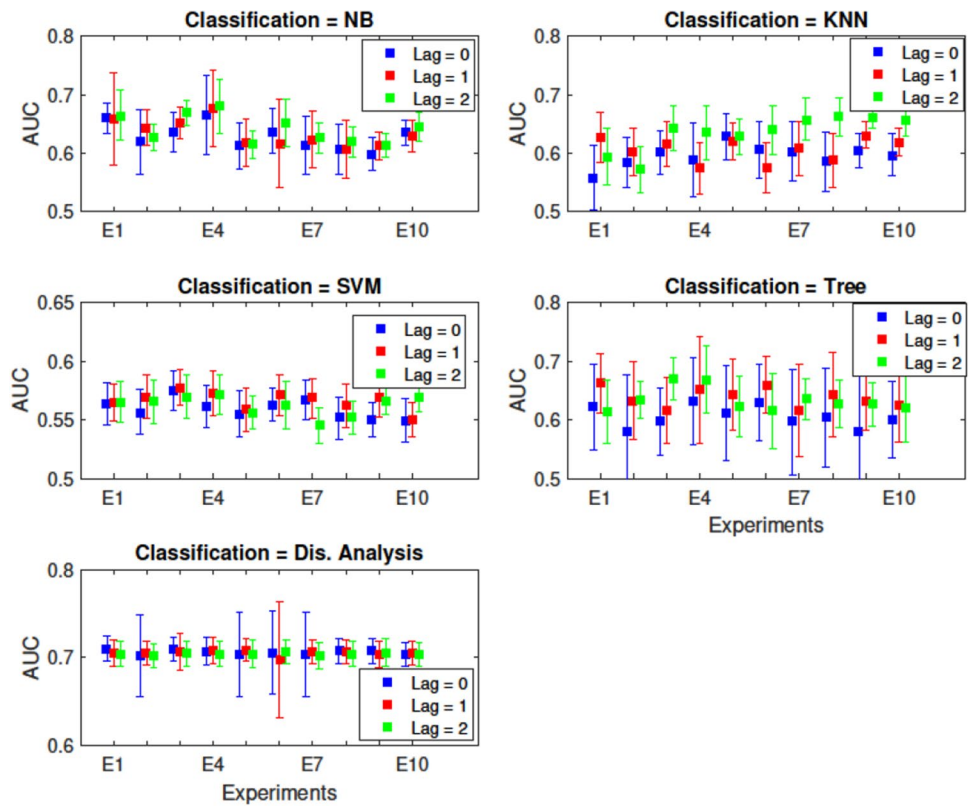
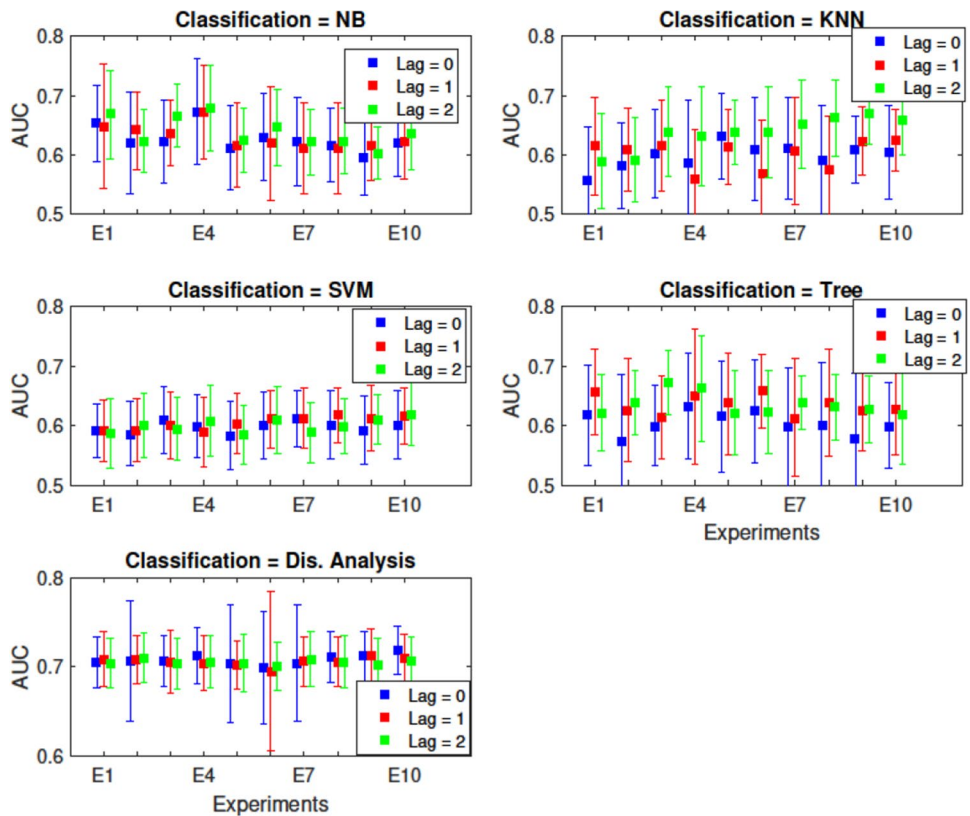


Fig. 8 Same as Fig. 7 but for test cases



different combinations of predictors with five classifiers were used in ten successive experiments for identifying the best choice. Each simulation was carried out using bootstrapping method, where 96 samples were randomly used to train the model and the rest of the samples were used for model validation. The bootstrapping procedure was applied for 100 times, and the ensemble means of model predictions of training and testing data were evaluated using ROC having 1 – specificity and sensitivity. The measure of model performance was AUC. Mean AUCs during the training and testing of all the experiments were compared and it was noted that DAC was the best model for predicting spring high-flows. The best predictor was the combination of $R_a + EC_a$ at lag = 0 having mean AUC of 0.72 (± 0.02) during model testing.

This study is among the very few efforts where JT-DP type spring high flows of the Indian Himalayas are predicted using multiple hydro-geological factors. Results of this study indicated that the long-term monitoring of rainfall and electrical conductivity of spring water alone would be sufficient to assess persistence of the JT-DP type spring high-flows in the central Himalayas. However, it is to be emphasized that the model performance could be improved by enhancing the temporal resolution as well as duration of the observed data. Similarly, model performance enhancement could further be carried out by including data from a higher number of JT-DP type springs. Since, rainfall climatology changes substantially across the Indian Himalayas, and the properties of aquifers contributing to spring discharges can substantially change with changing geology, a thorough investigation of hydrodynamical and geological properties of springs is extremely imperative before developing data derived model for a selected catchment.

Acknowledgements Research funding of NMHS, MoEFCC, GoI, is acknowledged (NMHS-2017-18/MG-02/478) for completing this work. Mr. Vinod Kanwal is highly acknowledged for collecting monthly spring discharge data and water samples. The Central Laboratory of GBPNIHE, Uttarakhand, India, is acknowledged for chemical analysis of the water samples. Mr. Sourab Singh and Kunal Joshi are acknowledged for preparing the spring land cover maps and geological information. The Director of GBPNIHE, Uttarakhand, India, is acknowledged for providing the computational facilities.

Author contributions Concept, data analyses, and writing: SM; concept and writing: SS; supervision and editing: KK. All authors agree with this version of the manuscript.

Funding Research funding of NMHS, MoEFCC, GoI, is acknowledged (NMHS-2017-18/MG-02/478).

Data availability The spring discharge and chemistry data are available with SM. The data sharing needs approval from the competent authority of GBPNIHE, Almora, India.

Code availability The classifier codes are available with SM. The code sharing needs approval from the competent authority of GBPNIHE, Almora, India.

Declarations

Conflict of interest No competing interests are present among the authors.

Ethics approval Not Applicable.

Consent for publication All the authors have their consents for publication.

References

- Agarwal A, Bhatnagar N, Nema RK, Agrawal N (2012) Rainfall dependence of springs in the midwestern Himalayan hills of Uttarakhand. *Mount Res Dev* 32:446. <https://doi.org/10.1659/MRD-JOURNAL-D-12-00054.1>
- Alfaro C, Wallace M (1994) Origin and classification of springs and historical review with current applications. *Environ Geol* 24(2):112–124. <https://doi.org/10.1007/BF00767884>
- Ballav S, Mukherjee S, Gosavi V, Dimri AP (2021) Projected changes in winter-season wet days over the Himalayan region during 2020–2099. *Theor Appl Climatol* 146(3–4):883–895. <https://doi.org/10.1007/s00704-021-03765-z>
- Banerjee A, Chen R, Meadows M, Singh RB, Mal S, Sengupta D (2020) An Analysis of long-term rainfall trends and variability in the Uttarakhand Himalaya using Google Earth Engine. *Remote Sens* 12(4):709. <https://doi.org/10.3390/rs12040709>
- Berrari D (2019) Bayes' theorem and naive Bayes classifier. In: Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics, vol 01. Elsevier, Amsterdam
- BIS 10500 (2012) Bureau of Indian Standards drinking water specification, second revision. Manak Bhawan, New Delhi, pp 16
- Brown RM, McClelland MJ, Deininger RA, O'Connor MF (1972) A water quality index—crossing the psychological barrier. In: Proceedings in international conference on water pollution research, pp 787–797
- Cheng S, Qiao X, Shi Y, Wang D (2020) Comparison of machine learning methods for predicting Karst spring discharge in North China. *ArXiv Preprint*. arXiv: 2007.12951
- Daniel D, Anandhi A, Sen S (2021) Conceptual model for the vulnerability assessment of springs in the Indian Himalayas. *Climate* 9(8):121. <https://doi.org/10.3390/cli9080121>
- Dass B, Abhishek SS, Bamola V, Sharma A, Sen D (2021) Assessment of spring flows in Indian Himalayan micro-watersheds—a hydro-geological approach. *J Hydrol* 598:126354. <https://doi.org/10.1016/j.jhydrol.2021.126354>
- Fawcett T (2006) Introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fiorillo F, Doglioni A (2010) The relation between Karst spring discharge and rainfall by the cross-correlation analysis. *Hydrogeol J* 18:1881–1895. <https://doi.org/10.1007/s10040-010-0666-1>
- Gebrehiwot A, Tadesse N, Jigar E (2011) Application of water quality index to assess suitability of groundwater quality for drinking purposes in Hantebet watershed, Tigray, Northern Ethiopia. *J Food Agri Sci* 1(1):22–30
- Granata F, Saroli M, de Marinis G, Gargano R (2018) Machine learning models for spring discharge forecasting. *Geofluids*. <https://doi.org/10.1155/2018/8328167>
- Gutierrez DD (2015) Machine learning and data science: an introduction to statistical learning methods with R. Technics Publications. <https://books.google.co.in/books?id=uAAcJgEACAAJ>

- Hu C, Hao Y, Yeh T-C, Pang B, Wu Z (2008) Simulation of spring flows from a Karst aquifer with an artificial neural network. *Hydrol Proc* 22:596–604. <https://doi.org/10.1002/hyp.6625>
- Kresic N, Bonacci O (2010) Spring discharge hydrograph. Groundwater hydrology of springs. Butterworth-Heinemann, Boston, pp 129–163
- Kumar K, Tiwari A, Mukherjee S, Agnihotri V, Verma RK (2019) Water at a glance: Uttarakhand. GBPNiHE, Almora
- Lambrakis N, Andreou AS, Polydoropoulos P, Georgopoulos E, Bountis T (2000) Nonlinear analysis and forecasting of a brackish Karstic spring. *Water Res Res* 36(4):875–884. <https://doi.org/10.1029/1999WR900353>
- Mayaud C, Wagner T, Benischke R, Birk S (2014) Single event time series analysis in a binary Karst catchment evaluated using a groundwater model (Lurbach system, Austria). *J Hydrol* 511:628–639
- Meinzer OE (1927) Large springs in the United States. USGS Water Supply Paper: Washington, USA
- Moriassi D, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Am Soc Agric Biol Eng* 50(3):885–900
- Mukherjee S (2021) Nonlinear recurrence quantification of the monsoon-season heavy rainy-days over northwest Himalaya for the baseline and future periods. *Sci Tot Environ* 789:147754. <https://doi.org/10.1016/j.scitotenv.2021.147754>
- Mukherjee S, Joshi R, Prasad RC, Vishvakarma SCR, Kumar K (2015) Summer monsoon rainfall trends in the Indian Himalayan region. *Theor Appl Climatol* 121(3–4):789–802. <https://doi.org/10.1007/s00704-014-1273-1>
- Mukherjee S, Ballav S, Soni S, Kumar K, De UK (2016) Investigation of dominant modes of monsoon ISO in the northwest and eastern Himalayan region. *Theor Appl Climatol* 125(3–4):489–498. <https://doi.org/10.1007/s00704-015-1512-0>
- Mukherjee S, Hazra A, Kumar K, Nandi SK, Dhyani PP (2019) Simulated projection of ISMR over Indian Himalayan region: Assessment from CSIRO-CORDEX South Asia experiments. *Meteorol Atmos Phys* 131(1):63–79. <https://doi.org/10.1007/s00703-017-0547-4>
- Murakami Y, Mizuguchi K (2010) Applying the naïve Bayes classifier with kernel density estimation to the prediction of protein–protein interaction sites. *Bioinformatics* 26(15):1841–1848. <https://doi.org/10.1093/bioinformatics/btq302>
- Namous M, Hssaisoune M, Pradhan B, Lee C-W, Alamri A, Elaloui A, Edahbi M, Krimissa S, Eloudi H, Ouayah M, Elhimer H, Tagma T (2021) Spatial prediction of groundwater potentiality in large semi-arid and Karstic mountainous region using machine learning models. *Water* 13(16):2273. <https://doi.org/10.3390/w13162273>
- Negi GCS, Joshi V (1996) Geo-hydrology of springs in a mountain watershed: the need for problem solving research. *Curr Sci* 71(10):772–776
- Negi GCS, Joshi V (2004) Rainfall and spring discharge patterns in two small drainage catchments in the Western Himalayan mountains, India. *Environmentalist* 24:19–28
- Negnevitsky M (2001) Artificial intelligence: a guide to intelligent systems. Pearson Education, Rotherham
- NITI Aayog (2018) Report of Working Group II sustainable tourism in the Indian Himalayan region. NITI Aayog, Government of India, Delhi, India
- Panwar S (2020) Vulnerability of Himalayan springs to climate change and anthropogenic impact: a review. *J Mt Sci* 17(1):117–132. <https://doi.org/10.1007/s11629-018-5308-4>
- Rani M, Joshi H, Kumar K, Bhatt DK, Kumar P (2021) Climate change scenario of hydro-chemical analysis and mapping spatio-temporal changes in water chemistry of water springs in Kumaun Himalaya. *Environ Dev Sustain* 23(3):4659–4674. <https://doi.org/10.1007/s10668-020-00793-z>
- Rathod R, Kumar M, Mukherji A, Sikka A, Satapathy KK, Mishra A, Goel S, Khan M (2021) Resource book on springshed management in the Indian Himalayan region: Guidelines for policy makers and development practitioners. International Water Management Institute (IWMI), India, New Delhi
- Savas C, Dovis F (2019) The impact of different kernel functions on the performance of scintillation detection based on support vector machines. *Sensors* 19(23):5219. <https://doi.org/10.3390/s19235219>
- Sifaou H, Kammoun A, Alouini MS (2020) High-dimensional linear discriminant analysis classifier for spiked covariance model. *J Mach Learn Res* 21:1–24
- Springer AE, Stevens LE, Anderson DE, Parnel RA, Kreamer DA, Levin L, Flora S (2008) A comprehensive springs classification system: integrating geomorphic, hydrogeochemical, and ecological criteria. In: Stevens LE, Meretsky VJ (eds) Arid land springs in North America: ecology and conservation. University of Arizona Press, Tucson
- Tambe S, Kharel G, Arrawatia ML, Kulkarni H, Mahamuni K, Ganeriwala AK (2012) Reviving dying springs: climate change adaptation experiments from the Sikkim Himalaya. *Mt Res Dev* 32(1):62–72. <https://doi.org/10.1659/MRD-JOURNAL-D-11-00079.1>
- Tarafdar S (2013) Understanding the dynamics of high and low spring flow: a key to managing the water resources in a small urbanized hill-slope of Lesser Himalaya, India. *Environ Earth Sci* 70(5):2107–2114. <https://doi.org/10.1659/10.1007/s12665-011-1493-y>
- Valdiya K, Bartarya S (1989) Diminishing discharges of mountain springs in a part of Kumaun Himalaya. *Curr Sci* 58:417–426
- White W (2010) Springwater geochemistry. In: Groundwater hydrology of springs, pp 231–268. <https://doi.org/10.1016/B978-1-85617-502-9.00006-2>
- Wunsch A, Liesch T, Cinkus G, Ravbar N, Chen Z, Mazzilli N, Jourde H, Goldscheider N (2021) Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrol Earth Syst Sci*. <https://doi.org/10.5194/hess-2021-403>
- Zhang Z (2016) Introduction to machine learning: K-nearest neighbors. *Ann Transl Med* 4(11):218–218. <https://doi.org/10.21037/atm.2016.03.37>
- Zhang YK, Bai EW, Libra R, Rowden R, Liu H (1996) Simulation of spring discharge from a limestone aquifer in Iowa, USA. *Hydrogeol J* 4(4):41–54

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.