# Cloud-based neuro-fuzzy hydro-climatic model for water quality assessment under uncertainty and sensitivity

Ananya Jain[1] · Srinivas Rallapalli[1,2] 🄳 · Dhruv Kumar[3]

## Abstract

River water quality is a function of various bio-physicochemical parameters which can be aggregated for calculating the Water Quality Index (WQI). However, it is challenging to model the nonlinearity and uncertain behavior of these parameters. When data is deficient and noisy, it creates missing and conflicting parameters within their complex inter-relationships. It is also essential to model how climatic variations and river discharge affect water quality. The present study proposes a cloud-based efficient and resourceful machine learning (ML) modeling framework using an artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS), and advanced particle swarm optimization (PSO). The framework assesses the sensitivity of five critical water quality parameters namely biochemical oxygen demand (BOD), dissolved oxygen (DO), pH, temperature, and total coliform toward WQI of the River Ganges in India. Monthly datasets of these parameters, river flow, and climate components (rainfall and temperature) for a nine-year (2011–2019) period have been used to build the models. We also propose collecting the data by placing various monitoring sensors in the river and sending the data to the cloud for analysis. This helps in continuous monitoring and analysis. Results indicate that ANN and ANFIS capture the nonlinearity in the relationship among water quality parameters with a root mean square error (RMSE) of $7.5 \times 10^{-7}$ (0.002%) and $1.02 \times 10^{-5}$ (0.029%), respectively, while the combined ANN-PSO model gives normalized mean square error (NMSE) of 0.0024. The study demonstrates the role of cloud-based machine learning in developing watershed protection and restoration strategies by analyzing the sensitivity of individual water quality parameters while predicting water quality under changing climate and river discharge.

**Keywords** Climate · Machine learning · Neuro-fuzzy · River discharge · Water quality

## Introduction

Unscientific utilization of water resources has resulted in its scarcity and contamination across different parts of the globe (Srinivas et al. 2018), especially in developing nations. Recent statistics indicate that riverine ecosystems of the largest rivers of the world such as Ganges (India),

Mississippi (USA), Jordan (Israel), Sarno (Italy), and Yellow (China) are severely affected due to agricultural and domestic sewages, multi-purpose projects, dams, and industrial manufacturing processes (Srinivas et al. 2017; Dwivedi et al. 2018). Hence, planners and scientific bodies need a robust approach to evaluate the water quality (Tripathi and Singal 2019; Srinivas et al. 2020b).

Water Quality Index (WQI) has been extensively used to determine the appropriateness of water for various beneficial purposes (Shah and Joshi 2017; Shil et al. 2019). Zotou et al. (2020) applied 7 different WQI methods to examine their application in water bodies of the Mediterranean region. Conventional WQI tools and other techniques such as multiple linear and nonlinear regression and auto-regressive aggregated moving average are unable to (i) capture the nonlinear and non-stationary aspects of water quality; (ii) analyze the sensitivity of individual water quality parameters; (iii) assess uncertain and subjective behavior of parameters;

Responsible Editor: Xianliang Yi

✉ Srinivas Rallapalli
   r.srinivas@pilani.bits-pilani.ac.in

1 Department of Civil Engineering, Birla Institute of Technology and Science, Pilani, Rajasthan, India

2 Department of Bioproducts and Biosystems Engineering, University of Minnesota, Twin Cities, Minneapolis, USA

3 Department of Computer Science and Engineering, University of Minnesota, Twin Cities, Minneapolis, USA

(iv) consider complex inter-relationships among parameters; and (v) develop a relationship with climatic parameters (Melesse et al. 2011; Oladipo et al. 2021). To overcome these shortcomings, researchers have also used artificial intelligence (AI)-based approaches such as fuzzy logic (Jha et al. 2020; Srinivas et al. 2017; Srinivas and Singh 2018a), neural networks (García-Alba et al. 2019; Ucun Ozel et al. 2020), neuro-fuzzy models (Aghaarabi et al. 2017; Azad et al. 2019), genetic programming (Sotomayor et al. 2018), and different hybrid models (Yaseen et al. 2018; Lu and Ma 2020). Several models such as PTMApp, ACPF, and HSPF-SAM have been coupled with AI techniques to assess the impact of various pollutant sources on impaired waters of watersheds in the USA (Srinivas et al. 2020a). Machine learning (ML) models such as adaptive neuro-fuzzy inference system (ANFIS) and hybrid ANN models can capture the nonlinear patterns of water quality by considering each parameter's sensitivity and subjective behavior to develop a robust WQI forecasting model (Tabbussum and Dar 2021). In addition, these models can also tackle problems such as high errors, outliers, noise, and missing data.

Kadam et al. (2019) projected the suitability of groundwater from the river basin Shivaganga to fulfill drinking purposes. Matta et al. (2020) classified spatio-temporal variation of water quality of the River Ganges (India) using WQI, multivariate statistical models, and principal component analysis (PCA). Azad et al. (2019) improved water quality classification and prediction by integrating artificial neural networks with nature-inspired algorithms like Cuckoo search, particle swarm optimization (PSO), and genetic algorithm. Juan et al. (2017) used ANN to simulate and forecast the variation of runoff in the Three-River Headwater Region (TRHR), Qinghai-Tibet Plateau. Al-Mukhtar and Al-Yaseen (2019) used ANN, ANFIS, and MLR to analyze electrical conductivity (EC) and total dissolved solids (TDS) in southern Iraq. Results established the supremacy of ANFIS among the three techniques. Fu et al. (2020) proposed an optimized wavelet de-noising ANFIS for estimating the water quality of wastewater. Pramanik and Panda (2009) compared ANN and ANFIS using daily data released from the Hirakud Reservoir, India. Sahu et al. (2011) demonstrated the capabilities of ANFIS to model actual and predicted water quality. They dealt with the data uncertainty and impreciseness by fuzzifying the parameter values.

Alizamir and Sobhanardakani (2018) used the ANN-PSO approach to predict heavy metals such as arsenic, copper, lead, and zinc in the groundwater of Toyserkan Plain (Hamedan Province, Iran). Chen and Liu (2015) simulated DO, phosphorus, chlorophyll 'a', and Secchi depth using ANFIS, radial basis function network, and a multiple linear regression model for Mingder reservoir (Taiwan). Aghel et al. (2019) modeled and predicted water quality parameters and achieved high computational speed and accuracy

to predict unknown parameters using a hybrid PSO–neural-fuzzy technique. Azad et al. (2019) enhanced the performance of ANFIS using GA, differential evolution, and ant colony optimization for continuous domains and determined water quality parameters of river Gorganroud along with sensitivity analysis. Zanganeh (2020) employed PSO algorithms to address some of the limitations of ANFIS and enhanced its accuracy.

Rivers are the most vulnerable of all ecosystems to climate change's effects. Hence, it is imperative to evaluate the role of climate variability on water quality (Avand and Moradi 2020). Variations in air temperature and rainfall are expected to have a significant impact on river flows, and as a result, they impact the mobility and dilution of pollutants as well (Srinivas et al. 2020c). Increased water temperatures can have an impact on chemical kinetics, which would impair the freshwater ecology and overall water quality (Mujere and Moyce 2018; Laanaya et al. 2017). Precipitation affects river flows, and as a result, sediment loads would vary, potentially altering river morphology and sediment transfer to water bodies, affecting freshwater ecosystems (Stryker et al. 2018). Avand and Moradi (2020) used integrated machine learning models with remote sensing to explore the impact of climate change and land use on water quality and flood probability. Hence, studying the uncertain behavior of climate parameters on river water quality is essential.

Recently, cloud computing is increasingly becoming popular to perform water quality analysis due to the increase in the computational requirements of analytical models such as machine learning (Geetha and Gouthami 2016; Sagan et al. 2020). The computational requirements may not necessarily be satisfied by locally available compute resources. To overcome the computational constraints, cloud computing platforms such as Amazon Web Services, Google Cloud Platform, and Microsoft Azure Cloud can be utilized for large-scale water quality analysis. The users can request for the required amount and type of computational resources on these cloud platforms. They will be charged with a monetary fee in turn for utilizing the cloud platform.

There are very few artificial intelligence-based models which simultaneously deal with water quality, climate change, and river flow along with sensitivity analysis of water quality parameters. Considering this, the present study proposes a cloud-based optimized hydro-climatic ML-based modeling framework using ANN, ANFIS, and ANN coupled with advanced PSO. Five critical water quality parameters namely BOD, DO, temperature, pH, and total coliform are used for predicting the WQI of the River Ganges (India) under changing climate. This study is aimed at.

(i). analyzing the sensitivity of water quality parameters toward predicting WQI using ANN, ANFIS, and ANN-PSO;

(ii). assessing the performance of proposed models using $R^2$, RMSE, and NMSE;

(iii). incorporating uncertain behavior of climate parameters (air temperature and precipitation) and river flow as well as their impact on water quality parameters using ANN and ANFIS;

(iv). comparing the modeling efficiency of ANN and ANFIS to predict WQI and assess the impact of hydro-climatology; and

(v). enabling all the above-mentioned analyses in the cloud by sending the data from the monitoring sensors placed in the river to the cloud.

Overall, the proposed advanced machine learning-based models would enable the watershed managers and decision-makers to predict water quality accurately in a continuous manner while incorporating uncertainties concerning climate change, water quality parameters, and river flow.

## Methodology

### Study area and data

The Ganges basin occupies around one-fourth (26.3%) of India's total area. River Ganges originates as Bhagirathi in the Garhwal Himalaya (30 55′ N, 79 7′ E). The catchment area of the Ganges basin is $8.6 \times 10^5$ km$^2$ (26.4% of India's area) (CPCB 2013). In Uttarakhand, the river traverses a length of 450 km. Raiwala is situated at an altitude of 356 m (above sea level) with an average rainfall of 2136.7 mm. Minimum and maximum summer temperatures are 29 ℃ and 40 ℃, respectively. In winters, the temperature drops

lowest to 5 ℃ and goes up to 20 ℃. The following lists the characteristics of the River Ganges in the study area (Raiwala district of Uttarakhand) (Fig. 1):

1) The downstream stretch of the River Ganges at Raiwala and Rishikesh is not suitable for bathing with respect to the BOD criteria.

2) Total coliform does not meet the primary water criteria based on the designated best use for category 'C' at d/s Raiwala and d/s Haridwar (CPCB 2013).

3) Drains in Rishikesh discharge 178.5 MLD directly to the Ganges River. Open defecation and discharge of untreated sewage (CPCB 2016) leads to higher concentrations of coliform bacteria and organic pollution.

The interaction of rising temperatures and changing discharge patterns due to climate change and the discharge of heavy chemicals and sewage into the river imbalances its ecosystem. Water quality has its unique relationship with climate variables (e.g., precipitation, temperature, daylight hours, wind speed). On the other hand, river discharge also plays a key role in regulating riverine ecosystems and overall self-purifying capacity (Srinivas et al. 2018b). Limited research has been performed to explore the relationships among climate change, river discharge, and water quality (Mimikou et al. 2000; Zhang et al. 2015). In this study, monthly datasets of water quality, climate parameters, and river discharge are collected for the time period of nine years, i.e., 2011–2019. Water quality parameters data (pH, BOD, DO, water temperature, and total coliform) at the Raiwala station are obtained from the Central Pollution Control Board (CPCB 2019). River discharge data is provided by Central Water Commission (CWC



**Fig. 1** Map of Ganges River basin highlighting the study area
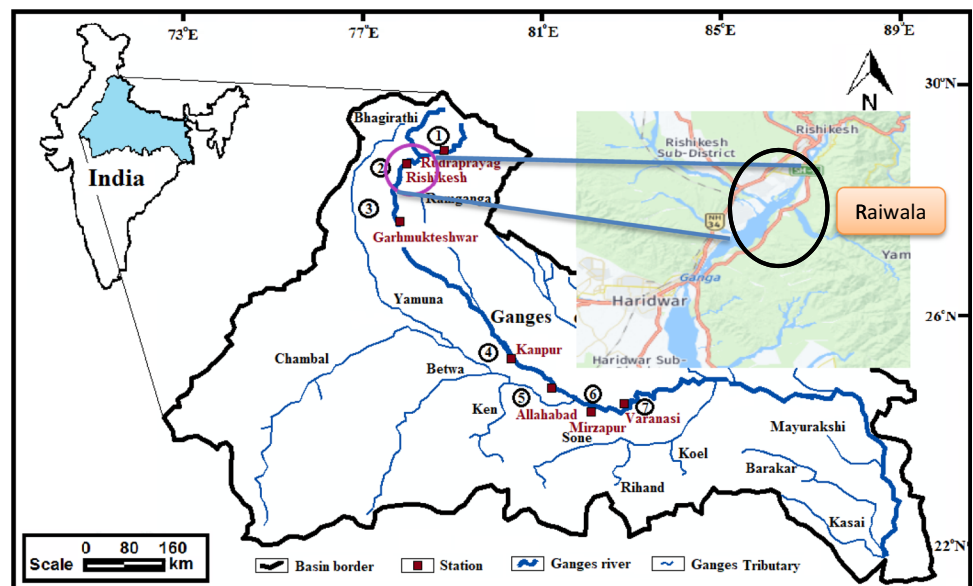
2019), while the climate data (rainfall and river temperature) is procured from Indian Meteorological Department (IMD 2019). For illustration purposes, water quality, climate parameters, and river discharge values only for the year 2011 are presented (Fig. 2). On analyzing WQI values for the study period (2011–2019), it is clear that the water quality of the river has significantly degraded from 2011 to 2019 primarily due to increased concentrations of total coliform and BOD.

## Optimized artificial intelligence models

The overall goal of the study is to accurately assess the impact of variation in hydro-climatic conditions and river discharge on water quality using ANN, ANFIS, and ANN-PSO. The methodology used in this study has been described in Fig. 3. We also propose collecting all the data from the monitoring sensors in the river and sending them to the cloud where the aforementioned analysis

**Fig. 2** Observed data of water quality, climate parameters, river discharge, and WQI at Raiwala (2011)
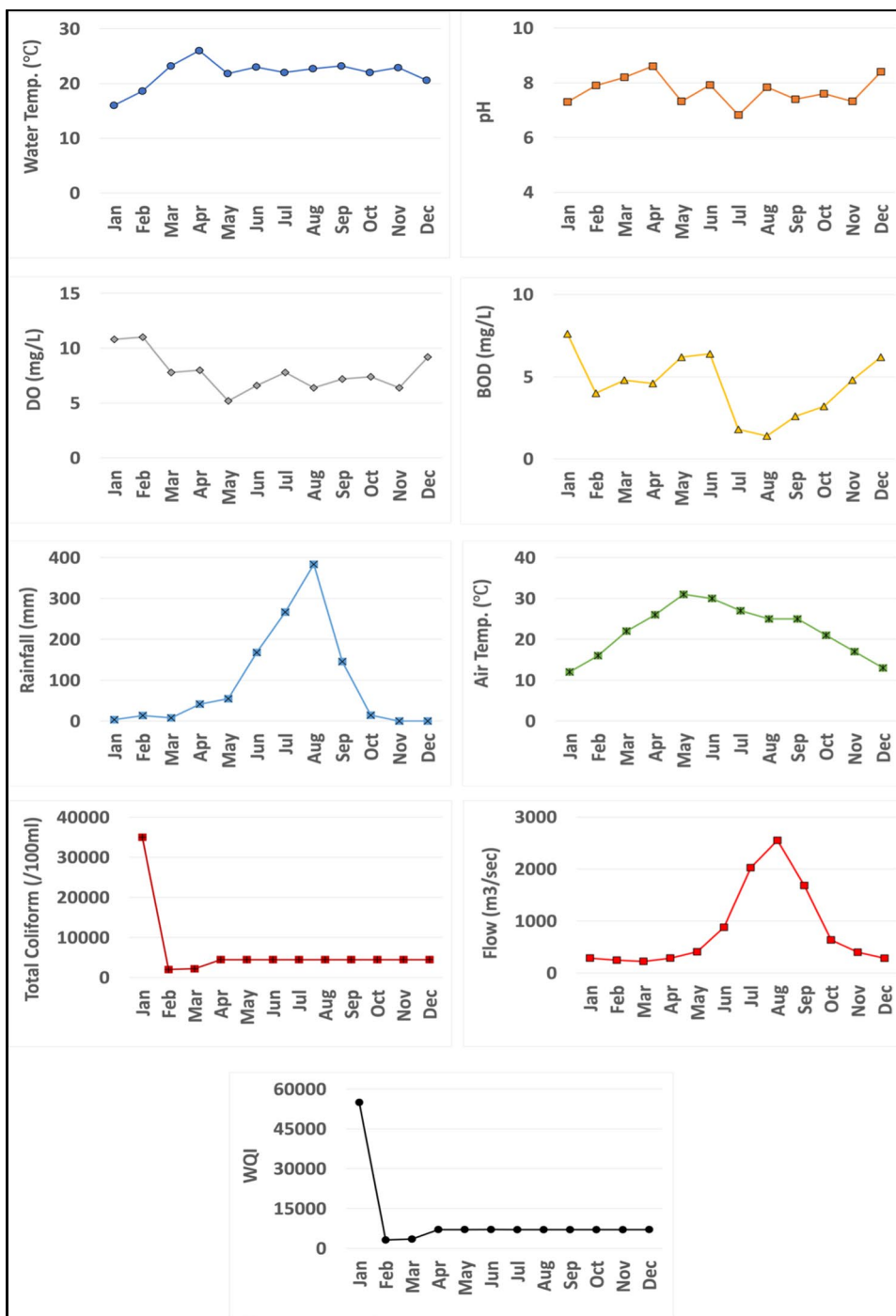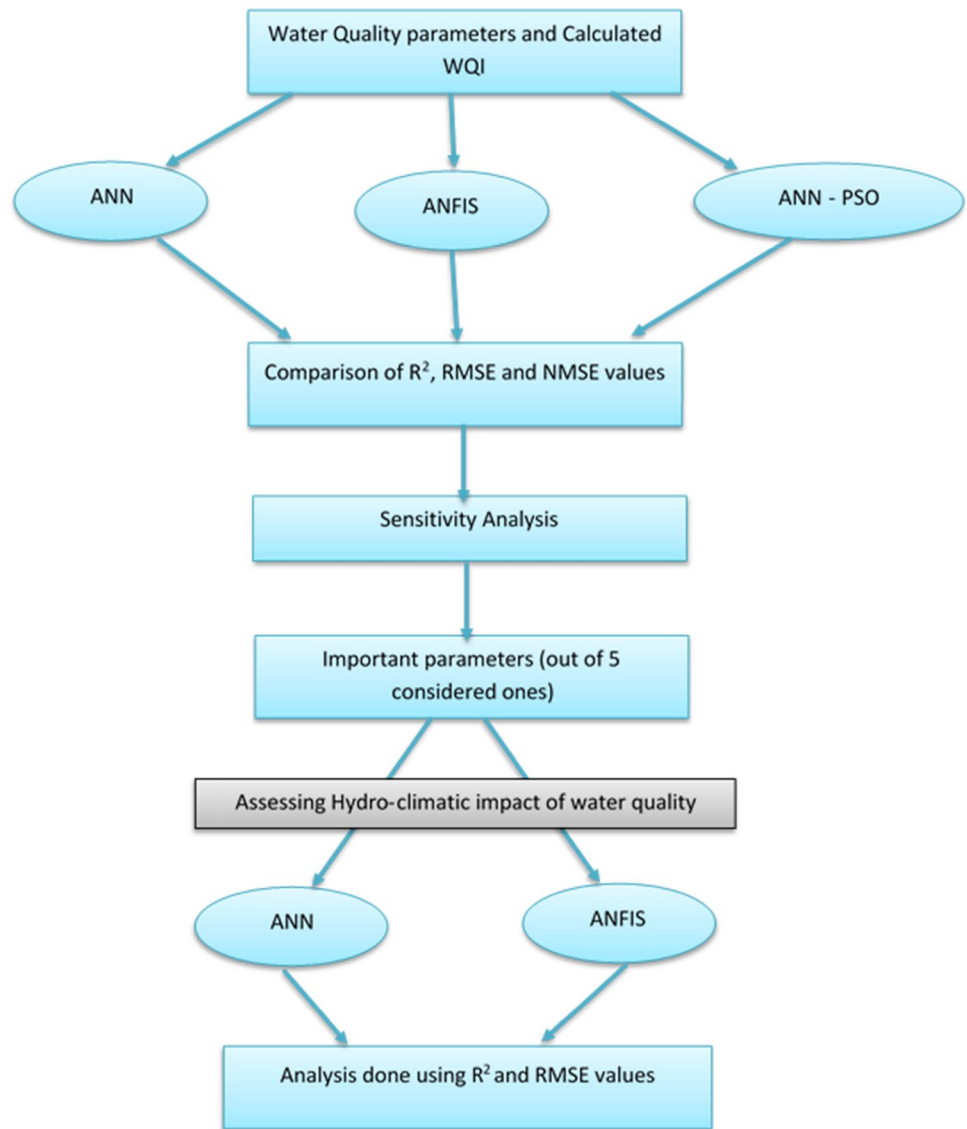
**Fig. 3** Methodology adopted in this study



will be carried out; this is represented in Fig. 4. Firstly, a sensitivity analysis of five water quality parameters with respect to WQI of River Ganges has been performed. These results are analyzed on the basis of root mean square errors (RMSE), normalized mean square error (NMSE), and the determination coefficient ($R^2$) for every combination of data using equations. Based on sensitivity analysis, critical parameters are obtained. Then, the impact of hydro-climatology on water quality is accessed using ANN and ANFIS. The models are applied to the case study on River Ganges in the following manner:

WQI calculation: Firstly, monthly WQI values are calculated using the weighted arithmetic mean method for the 9-year monthly data (Sutadian et al. 2016) using Eqs. (1)–(3).
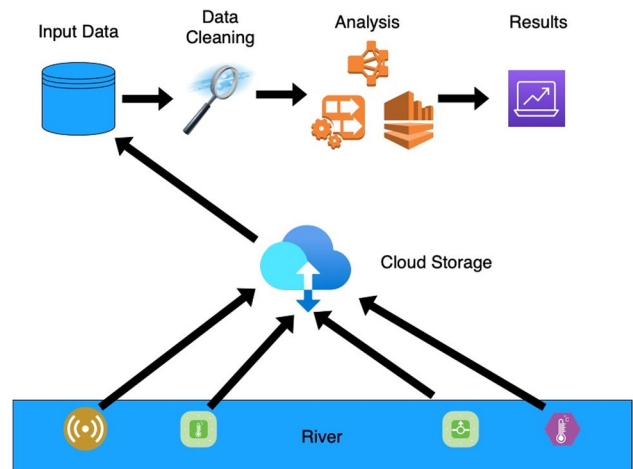


**Fig. 4** Utilizing the cloud for continuous monitoring and analysis

$$q_n = \frac{100(X_n - X_{io})}{(P_n - X_{io})} \tag{1}$$

$$W_n = {K}/{P_n} \tag{2}$$

$$WQI = \frac{\sum q_n W_n}{\sum W_n} \tag{3}$$

where $n$ is the number of water quality parameters, $q_n$ is the qualitative rating of the $n$th parameter, $X_n$ is the observed value of the $n$th parameter at the sampling station, $P_n$ is the standard prescribed value of the $n$th parameter, $V_{io}$ is the ideal value of the $n$th parameter in pure water, i.e., zero for all other parameters except pH and DO (7.0 mg/L and 14.6 mg/L, respectively), $K$ is the constant for proportionality, and $W_n$ is the unit weight of the $n$th parameter. Table 1 provides the drinking water standards and unit weights of parameters provided by the Bureau of Indian Standards (BIS) and the World Health Organization (WHO).

Normalization: The entire input datasets (water quality parameters) are normalized on a scale of 0 to 1 (Srinivas et al. 2017).where $f$ is the output of the model, i.e., WQI and important water quality parameters in our study, $p = 1,2,3…….N$ and $N$ is the number of data samples. Identifying sensitive water quality parameters: The normalized values of five water quality parameters and WQI are considered as input and output, respectively, to each of the ANN, ANFIS, and ANN-PSO models. Models are trained using different combinations of four water quality parameters at a time. Additionally, models are also trained by feeding all five water quality parameters as input. By using statistical methods ($R^2$, RMSE, and NMSE) [Eqs. (4)–(6)], sensitive parameters are obtained for further analysis.where $f$ is the output of the model, i.e., WQI and important water quality parameters in our study, $p = 1,2,3…….N$ and $N$ is the number of data samples.

**Table 1** Drinking water standards recommending agencies and unit weights

| Parameter | Standard value | Unit weight |
|---|---|---|
| pH | 6.5–8.5 | 0.2190 |
| DO (mg/l) | 5 | 0.3723 |
| BOD (mg/l) | 5 | 0.3723 |
| Temp. (℃) | 25–50 | 0.0372 |
| Total coliform (/100 ml) | 10 | 0.1861 |

$$R^2 = \frac{\sum_{k=1}^{N} \left[ \left( f(p) - \overline{f(p)} \right) \left( f(\hat{p}) - \overline{f(p)} \right) \right]}{\sqrt{\sum_{k=1}^{N} \left( f(p) - \overline{f(p)} \right)^2 \sum_{k=1}^{N} \left( f(\hat{p}) - \overline{f(\hat{p})} \right)^2}} \tag{4}$$

$$RMSE = \sqrt{\sum_{K=1}^{N} \frac{1}{N-1} [f(p) - f(\hat{p})]^2} \tag{5}$$

$$NMSE = \frac{\sum_{k=1}^{N} [f(p) - f(\hat{p})]^2}{\sum_{k=1}^{N} [f(p) - \overline{f(p)}]^2} \tag{6}$$

Assessing hydro-climatic impact: Another set of ANN and ANFIS models are trained for identifying the impact of climate and river discharge on water quality. A separate ANN and ANFIS models are built for each of the sensitive water quality parameters. Each model consists of flow rate of the River Ganges, air temperature, and rainfall at Raiwala as inputs while the water quality parameters as outputs. The trained ANN and ANFIS models are analyzed to assess the level of impact of climate and river flow on water quality using $R^2$ and RMSE.

## ANN and ANFIS models

In artificial neural networks (ANNs), an activation function (which is generally nonlinear) is applied to inputs (either water quality parameters or climate and discharge data) arriving at each neuron for generating an output signal (either WQI or water quality parameters). An additional bias input is added to the weighted sum of inputs to increase/decrease the overall input to the activation function. Such synthesized functions depend on the network architecture and interconnections between the processing units. Most ANN algorithms are feedforward networks which use error backpropagation algorithms for training. The input and output nodes represent independent and dependent variables, respectively. Hidden layers are added to perform nonlinear transformations (such as sigmoid functions) on the input space. Detailed information on ANN can be found in (Rajaee et al., 2020).

An adaptive neuro-fuzzy inference system (ANFIS) uses a fuzzy inference system (FIS) to express the uncertainty associated with water quality, river discharge, and climate parameters. FIS uses a nonlinear mapping from the input to the output space using a set of 'IF–THEN' fuzzy laws, each defining the mapping's local behavior. In general, backpropagation alone or backpropagation along with the least square approach is used to optimize the fuzzy membership parameters. ANFIS model is a stronger model than ANN and independent fuzzy logic models in terms of its flexibility, adaptivity, learning speed,

and memorization accuracy (Srisaeng et al., 2015). It is also able to better capture the nonlinearity, uncertainty, and inaccuracy in the data.

Sugeno-type FIS has been used in this study as it models the real-world data more accurately and is more computationally efficient than Mamdani (Wang and Chen 2014; Şahin and Erol 2017). The Sugeno system has either linear or constant output membership functions, whereas the Mamdani system can have triangular, Gaussian output membership functions. The Mamdani type is more reliant on expert knowledge and requires fuzzy rules. On the other hand, the Sugeno-type FIS utilizes real data for training, and it develops the rules itself while training according to the input and output data. In general, Sugeno FIS is used for one output and Mamdani FIS is used if the number of outputs is more than one. In this study, there is only one output, i.e., WQI, and the model needs to be trained based on real-time data; hence, Sugeno FIS is chosen. Figure 5 shows the architecture of an ANFIS model, where $O_{j,i}$ represents the output of the $i$th node and $j$th layer, $x$ is the input (water quality parameters) to the $i$th node, and Ai is the linguistic label of inputs denoted by membership functions (triangular, trapezoidal, and Gaussian).

The fuzzy IF–THEN rules for a first-order Sugeno fuzzy model are given below:

Rule 1 : If $x$ is $A_1$ and $y$ is $B_1$, then $f_1 = a_1x + b_1y + c_1$,

Rule 2 : If $x$ is $A_2$ and $y$ is $B_2$, then $f_2 = c_2x + b_2y + c_2$,

where $x$ and $y$ are inputs, $A_i$ and $B_i$ are membership functions, $f_i$ is the output, and $a_i$, $b_i$, and $c_i$ are determined during the training process (Şahin and Erol 2017). The layers of the ANFIS model are explained below.

Layer 1: Nodes are represented through the following functions:

$$O_{1,i} = \mu_{Ai}(x) \ for \ i = 1, 2$$

$$O_{1,i} = \mu_{Bi-2}(y) \ for \ i = 3, 4$$

Layer 2: Firing strength of a rule is computed as follows:

$$O_{2,i} = w_i = \mu_{Ai}(x) \times \mu_{Bi}(y) \ where \ i = 1, 2$$

Layer 3: Normalized firing strengths are calculated as

$$O_{3,i} = \overline{w}_1 = {w_i}/{w_1+w_2} \ i = 1, 2$$

Layer 4: The node function of this layer calculates the contribution of the $i$th rule to the total output

$$O_{4,i} = \overline{w_1} \times f_i = \overline{w_1}(a_i \times x + b_i \times y + c_i)$$

where $w_i$ is the output of layer 3 and $\{a_i, b_i, c_i\}$ is the parameter set.

Layer 5: Single output node computes the overall output of the ANFIS as

$$O_{5,i} = \sum_i \overline{w_1} \times f_i = {\Sigma_i w_i \times f_i}/{\Sigma_i w_i}$$

In this study, a backpropagation learning algorithm is used to reduce the error between observed and expected data. A well-designed ANFIS has been developed to address the nonlinearity or complexity associated with inputs and outputs with high precision.

### ANN training and architecture

*Feedforward backpropagation neural network*: A total of 108 normalized (on a scale of 0–1) input and output values are fed into the hidden layer, and desired weights are assigned. The number of hidden layers affects the error estimation. The weights connect each link between an input and a hidden layer, as well as a bias value. The output layer then connects to the hidden layer, which gives the output. All the

**Fig. 5** Architecture of ANFIS model having two inputs, two rules, and one output
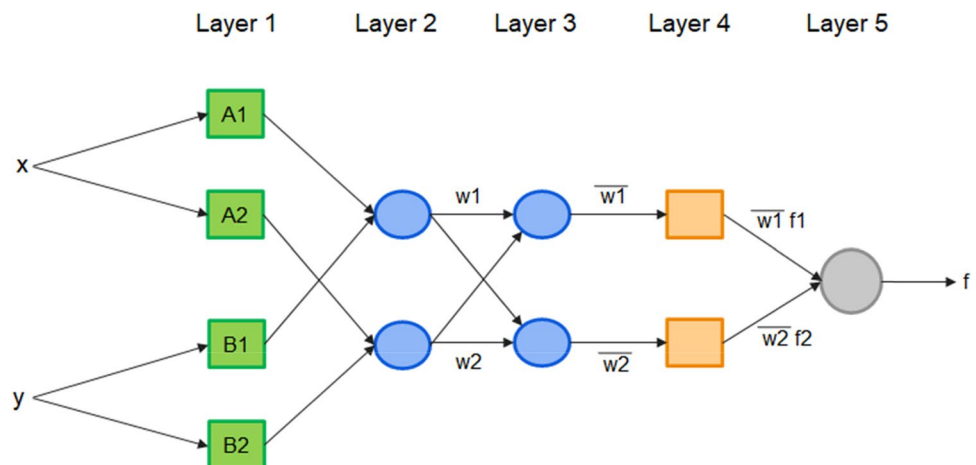
neural network models generated in this study are analyzed using $R^2$ and RMSE.

**Training the neural network** Training achieves highly efficient and accurate modeling of a given dataset. For sensitivity analysis, input data includes five water quality parameters and output includes the WQI of the river. For assessing the impact of hydro-climatic changes, river discharge, rainfall, and air temperature are considered as inputs; and the output dataset includes pH, DO, BOD, temperature, and total coliform. The backpropagation algorithm is used for the training. Using the fitnet function in MATLAB 2020a, Levenberg–Marquardt (LM) function has been chosen as the training algorithm and tansig as the activation function for hidden as well as output layers.

**Neural network architecture** The structural attributes of a feedforward neural network such as the number of hidden layers, hidden neurons, and training, testing, and validation datasets directly affect the performance of the model and are important for adequate mapping. The trial-and-error process has been used extensively to arrive at a perceptron with three layers consisting of a number of functional nodes and connection weights. Input and output values are randomly partitioned into 70%, 15%, and 15% as training, testing, and validation sets, respectively. Figure 6 represents the architecture of the ANN model for analyzing the impact of variation in hydro-climatology on water quality. Because of the lack of availability of a huge dataset, overfitting in the ANN model is dealt with by changing the network complexity. The network

structure was changed in a manner that the parameters (weights) of the model remain small. Small parameters suggest a less complex and, in turn, more stable model that is less sensitive to statistical fluctuations in the input data. A trial-and-error process was undertaken till the model achieved a suitable performance on a validation dataset, and the corresponding network structure was finalized.

Following the above-mentioned concept, while training an ANN model with five water quality parameters as input and Water Quality Index (WQI) as the output, taking one hidden layer between the input and output layer was the basic starting point so that number of weights in the model could be optimized to a minimum level. The model was developed with 15% of the dataset as a validation dataset. For deciding the number of hidden neurons in that layer, trial and error was done with 5, 6, and 7 neurons. The values of the regression coefficient for all the three models with 5, 6, and 7 hidden neurons were similar (around 0.99). This indicated that all these three models were able to train well. However, the predictive error on the validation dataset was the least for the model with 5 neurons with also the training error being the least ($7.5 \times 10^{-7}$). For 6 and 7 neurons, the predictive error of the validation dataset was very high as compared to the low training error. Therefore, with 5 input parameters, 5 hidden neurons in 1 hidden layer and 1 output node (WQI) in the output layer, the neural network model was developed to achieve, if any, the least amount of overfitting.

**Fig. 6** Architecture of ANN model for climate change analysis
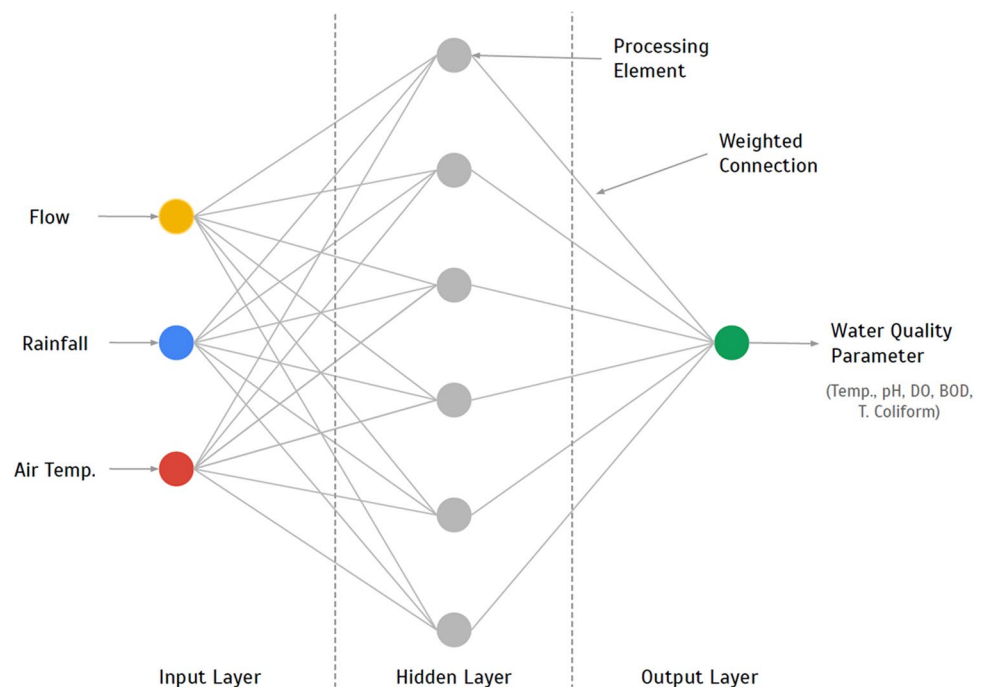
## ANFIS training and architecture

Both sensitivity and hydro-climatic analysis use a combination of input and output values similar to that of the feedforward backpropagation of ANN, as described earlier. Also, $R^2$ and RMSE are calculated.

**Sugeno-FIS architecture** The ANFIS model's ability to meet the output target is determined by internal ANFIS parameters including the number and shape of membership functions. Each input parameter's contribution to the regression parameter and the desired output is determined by the type and number of membership functions used. Optimization of internal parameters is therefore a necessary step during ANFIS modeling. If the target RMSE can be met, the trial-and-error method is considered to be efficient for such optimization. The trial-and-error method also has the benefit of producing an information rule base with a lower likelihood of overfitting the training data collection. Hence, internal ANFIS learning parameters are tuned to render the optimal parameters capable of mimicking the given data pattern sequences after a comprehensive trial-and-error process. For every input, three Gaussian membership functions are used. For illustration, Fig. 7 shows the membership functions used for pH in this study. The Gaussian membership function ($\mu_A$) is generally represented as Gaussian ($x$: $b$, $s$), where $b$ and $s$ represent the mean and standard deviation, respectively (Eq. (7)).

$$\mu_A(x, b, s, n) = \exp\left[-\frac{1}{2}\left|\frac{x - b}{n}\right|^n\right] \tag{7}$$

where $n$ represents the fuzzification factor.

The output membership function of the ANFIS model is linear. The number of inputs, their types, and the number of fuzzy membership functions used in the model are used to determine the number of fuzzy rules and the optimum number of parameters required for describing the FIS to obtain the best results.

**Training the ANFIS** Training is done with the aid of a neural-fuzzy designer in MATLAB 2020a. For illustration, the architecture of the ANFIS model for sensitivity analysis has been shown in Fig. 8. The data space is divided into rectangular subspaces using grid partition and axis-parallel partition on the basis of a predetermined number of membership functions (MFs) and their types in the dimensions. Backpropagation along with the least-squares algorithm (hybrid) adjusts the parameters of the membership functions. The parameters of input membership functions are adjusted using backpropagation, while those of output membership functions are adjusted using least-squares estimation in this hybrid optimization process. In this study, the training stops after 50 epochs as there is no change in error after 50 epochs.

### Hybridized particle swarm optimization (PSO)-ANN model

Meta-heuristic optimization algorithms are used to train ANNs instead of classic algorithms (backpropagation algorithm) to tackle the disadvantages associated with the latter like getting stuck in local minima/maxima and plateaus of the error function landscapes (Chatterjee et al. 2017). In a hybrid ANN-PSO model, PSO is used to reduce the ANN's errors by selecting the best values for the model's weights and biases. As a result, the weights and biases are variables in this problem, and the problem's feasible space is determined by the interval at which these variables differ.

In ANN-PSO, each particle contains one set of weights and the value of weights corresponding to a particle is termed as the particle's position. A particle's fitness function can be expressed in terms of NMSE. Firstly, a number of hidden layers of neurons are chosen, and a neural network is developed with initial weights and biases. The initial weights and biases are then improved in order to be able to locate the particle in the D-dimensional, where $D$ refers to the total number of weights and biases. In every iteration, each particle's output values are predicted, followed by the calculation of normal mean squared error (cost function in this study). The location of particles is updated for a specified number of populations and iterations until the cost function is

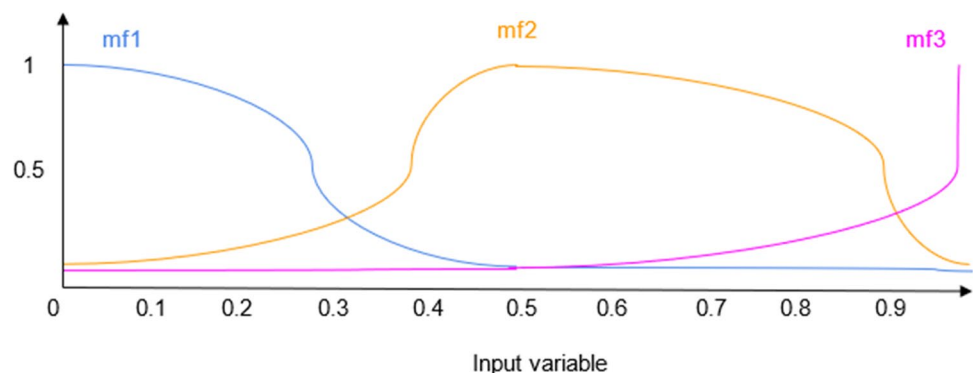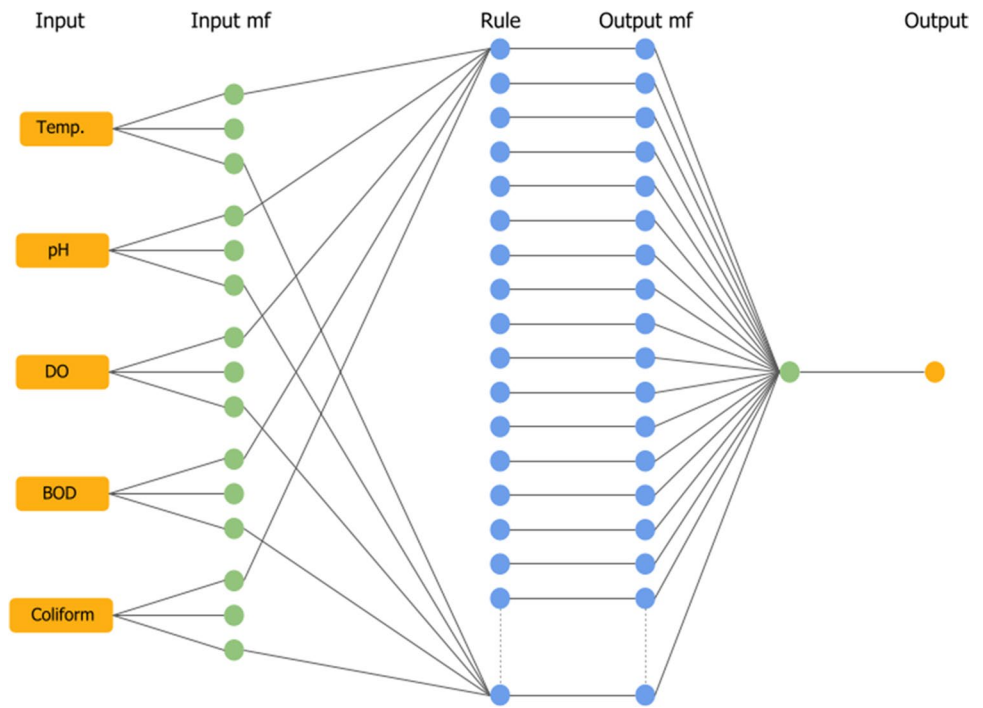**Fig. 7** Gaussian membership function plot for an input variable pH

**Fig. 8** Architecture of ANFIS model in Neuro-Fuzzy Designer for sensitivity analysis



minimized (Poli et al. 2007). For this study, the weights refer to the weights of the ANN model with water parameters as input and WQI as output for sensitivity analyses. A total of 50 particles (population) have been considered in the PSO algorithm. Out of these 50 particles, the most optimum one whose position when fixed as the weights in the ANN model gives the least NMSE after training.

In the proposed work, PSO along with controlled parameters has been proposed to improve the ANN model. To investigate particle movement patterns in PSO with controlled parameters, base frequency ($F$) and variance ($M_c$) are used to estimate particle position coefficients. The term 'base frequency' is used to describe the movement as a measure of positional correlation. The range of movement patterns is characterized by variance ($M_c$). The frequency and variance are measured theoretically, and the inertial weight ($w$) is calculated using Eqs. (8)–(10).

In comparison to a long-run problem, the number of iterations for a short-run problem is lower. If a short run and a long run are conducted, the experimental values of $M_c$ and $F$ are given as 25.6 and 0.2, respectively, for a short run, 6.4 and 0.2, respectively, for a long run (Bonyadi and Michalewicz, 2016).

### ANN-PSO architecture

Particle swarm optimization has been used to improve the ANN modeling of water quality parameters and WQI while performing sensitivity analysis. $R^2$ and NMSE are used to compare the models and evaluate the sensitivity of each water quality parameter.

**Neural network architecture** A feedforward perceptron with one hidden layer is improved using advanced PSO. The

$$f(w) = \frac{-48V_c w^2 + 48M_c}{28M_c + w - 20M_c w + 1} - 1 - w + 2 \cos(2\pi F)\sqrt{w} \, ; f(w) = 0 \quad (8)$$

For a given frequency and variance, values of coefficients are calculated as

$$c_1 = 1 + w - 2\cos(2\pi F)\sqrt{w} \quad (9)$$

$$c_2 = \frac{-48V_c w^2 + 48M_c}{28M_c + w - 20M_c w + 1} \quad (10)$$

training algorithm and activation functions are the same as in the 'ANN training and architecture' section. Input (pH, DO, BOD, temperature, and total coliform) and output (WQI) values are randomly divided into 70%, 15%, and 15% as training, testing, and validation sets, respectively.

**Optimization of weights** Obtaining suitable weights between 'input and hidden layers' and 'hidden and output

layers' using a meta-heuristic nature-inspired algorithm has proven to be extremely efficient as compared to classical methods. The number of weights to be optimized is calculated using Eq. (11).

$$\text{No. of weights} = (\text{No. } \textit{of } \text{input nodes} * n) + n + n + 1 \tag{11}$$

where $n$ is the number of hidden nodes.

In the PSO algorithm, values of these weights are considered as positions corresponding to each particle. Particle size is known as population. The velocity of each particle is also generated, and after each iteration, both the position and velocity of the particle are updated based on global best values. Following a trial-and-error method to get the best performance, iterations are fixed to a maximum of 200, the population is 50, and the damping coefficient is set to be 0.99. In order to enhance PSO, controlled parameters are used (Bonyadi and Michalewicz, 2016) as given below:

Inertial weight $(w) = 0.836$

Acceleration constants $(c_1$ and $c_2) = 1.271$

**Fitness function** The fitness (cost) function in the PSO algorithm equals NMSE. After being trained with PSO updated weights, ANN gives the error for a specific input and output dataset used for modeling the trained neural network. These error and output values are used to calculate the NMSE of those combinations of weights. The weights that produce the least value of the fitness function (NMSE) are chosen for modeling the ANN. Figure 9 shows how error decreases as the iterations progress to finally reach the least NMSE.

## Results and discussion

The AI models described in the previous sections have been applied using the nine-year (2011–2019) datasets of water quality and climate parameters, and discharge at the Raiwala sampling station of the River Ganges. This section presents the results for each of the ANN, ANFIS, and ANN-PSO models.

### Sensitivity analysis of water quality parameters

ANN, ANFIS, and ANN-PSO have been applied to carry out a sensitivity analysis to identify the water quality parameters which affect the WQI of the Ganges River in a significant manner. Table 4 presents the results of the ANN models. Based on the results, the combination of one hidden layer with five hidden neurons is selected, as it yields minimum RMSE ($7.50 \times 10^{-7}$). However, on selecting six or seven hidden neurons, RMSE increases. Therefore, five hidden neurons are selected for water quality parameters sensitivity analysis. Five models are trained, each time leaving one parameter out, using the fitnet function (nnstart toolbox) in MATLAB 2020a. On removing any parameter from the analysis, the RMSE and $R^2$ values of the neural models increase or decrease, signifying their importance accordingly. 'ANN-WQI-ALL' is taken as the reference model. When all five parameters are taken as input, a correlation coefficient of

**Fig. 9** Relation between iterations and NMSE for ANN-PSO model with five water quality inputs
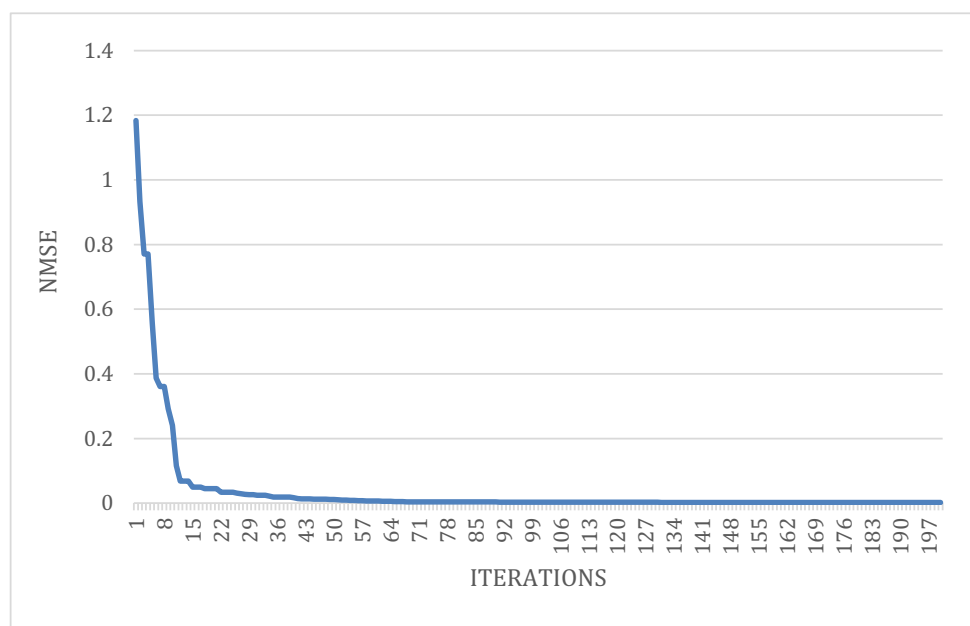
0.999 is obtained. On the other hand, if BOD is not taken as input, a correlation coefficient of 0.856 and RMSE of 0.0659 is obtained. Hence, comparing the values of statistical measures, it is observed that pH, BOD, and total coliform have the largest influence on the WQI. DO ($R^2 = 0.965$ and RMSE $= 0.044$) and water temperature ($R^2 = 0.999$ and RMSE $= 5.53 \times 10^{-6}$) have comparatively lesser impact on WQI (Table 2).

Table 3 presents the results of the ANFIS models. The ANFIS-WQI-ALL model has the lowest RMSE value ($1.02 \times 10^{-5}$) and shows good accuracy as well as little residual errors when compared to other models. Hence, it is taken as the reference model. RMSE increases greatly (0.000131) when total coliform is not taken into consideration. This reveals that total coliform can be considered as the most crucial parameter for WQI prediction. Low errors corresponding to 'ANFIS-WQI-BOD' and 'ANN-WQI-pH' also indicate the importance of BOD and pH as inputs when compared to temperature and DO. The $R^2$ values do not vary significantly from 0.99 indicates that ANFIS has been able to achieve very good training accuracy on the provided data.

Table 4 gives the results of sensitivity analysis for ANN-PSO models. On taking 7 hidden neurons, all parameters when considered into the model yield an $R^2$ value of 0.998 and NMSE of 0.0024. Optimized weights through PSO gave the best results when the ANN architecture had one hidden layer and 7 hidden nodes. The highest dip in $R^2$ value (0.7196) is seen in the case of ANN PSO-WQI-COLIFORM. Consequently, the NMSE value of 0.3103 proves that total coliform is the most significant input parameter. The remaining models are statistically close with one another.

Because of their ability to yield more accurate results, ANFIS and ANN-PSO are considered more reliable compared to ANN models. The observations obtained from the results of these models can be easily verified based on the works performed by governmental bodies and other secondary sources (CPCB 2016; Srinivas et al. 2017; UPCB 2020). In Rishikesh, a total of 187.92 MLD of wastewater

sewage and 1.27 tonnes/day of BOD is discharged into the River Ganges (CPCB 2016). On major religious days, the discharge of municipal sewage increases significantly as more than 1.5 million people gather near the River Ganges. In addition, Uttarakhand is the home of 71 grossly polluting industries discharging 94,992 KLD of industrial waste and 2150 kg/day of BOD into the River Ganges and its tributaries (ENVIS 2016). These industries include paper pulp, sugar, distilleries, and pesticides. Since most of them discharge BOD, pH at high temperatures, a strong correlation is obtained using ANN, ANFIS, and ANN-PSO models due to the presence of a common source. BOD, pH, and coliform are also present in sewage water; thus, coliform is also closely related to other parameters.

Due to constant discharge from sewage treatment plants, untreated drains, and industries across the Dehradun-Rishikesh stretch, BOD and coliform exceed the water quality criteria (CPCB 2019). pH levels increase tremendously, particularly in the monsoon seasons. Haritash and Gaur (2016) conclude that Rishikesh's water cannot be advised for drinking or other domestic needs unless it is treated.

This study would enable governmental bodies such as CPCB to take necessary precautions against the discharge of sewage, human, and animal waste into the waters of the River Ganges and devise suitable measures to assess the climatic impact on aquatic life. It would also help the

**Table 3** Overall results of six ANFIS-WQI models developed for sensitivity analysis

| Models ↓ | $R^2$ | RMSE |
|---|---|---|
| ANFIS-WQI-ALL | 0.999 | $1.02 \times 10^{-5}$ |
| ANFIS-WQI-TEMP | 0.985 | $2.43 \times 10^{-5}$ |
| ANFIS-WQI-PH | 0.973 | $3.23 \times 10^{-5}$ |
| ANFIS-WQI-DO | 0.989 | $2.26 \times 10^{-5}$ |
| ANFIS-WQI-BOD | 0.998 | $3.31 \times 10^{-5}$ |
| ANFIS-WQI-COLIFORM | 0.997 | 0.000131 |

**Table 2** Overall results of eight ANN-WQI models developed for sensitivity analysis

| Models ↓ | Hidden neurons | $R^2$ | RMSE |
|---|---|---|---|
| ANN-WQI-ALL | 5 | 0.999 | $7.50 \times 10^{-7}$ |
|  | 6 | 0.999 | $2.03 \times 10^{-5}$ |
|  | 7 | 0.999 | $1.75 \times 10^{-5}$ |
| ANN-WQI-TEMP | 5 | 0.999 | $5.53 \times 10^{-6}$ |
| ANN-WQI-PH | 5 | 0.844 | 0.066 |
| ANN-WQI-DO | 5 | 0.966 | 0.044 |
| ANN-WQI-BOD | 5 | 0.856 | 0.065 |
| ANN-WQI-COLIFORM | 5 | 0.827 | 0.056 |

**Table 4** Overall result of eight ANN PSO-WQI models developed for sensitivity analysis

| Model ↓ | Hidden neurons | $R^2$ | NMSE |
|---|---|---|---|
| ANN PSO-WQI-ALL | 7 | 0.9987 | 0.0024 |
|  | 6 | 0.9882 | 0.0121 |
|  | 5 | 0.9961 | 0.0039 |
| ANN PSO-WQI-TEMP | 7 | 0.9934 | 0.0068 |
| ANN PSO-WQI-PH | 7 | 0.9948 | 0.0052 |
| ANN PSO-WQI-DO | 7 | 0.9955 | 0.0045 |
| ANN PSO-WQI-BOD | 7 | 0.9934 | 0.0067 |
| ANN PSO-WQI-COLIFORM | 7 | 0.7196 | 0.3103 |

concerned authorities to plan and manage big pilgrimages and the establishment of industries and factories in a safe and efficient manner.

## Suitable models for WQI forecasting

ANN, ANFIS, and ANN-PSO show great capability during the training stage when used for the numerical modeling of WQI, considering pH, BOD, water temperature, and total coliform as inputs. However, ANFIS models are the most robust and accurate, with no major fluctuations in the values of $R^2$ and lesser residual errors (Table 3). ANFIS models are observed to be superior over ANN and hybrid ANN models (Kisi 2015). They have proven to efficiently deal with areas that are ill-defined and ambiguous, such as water quality projections (Aghaarabi et al. 2017; Tiwari et al. 2018) and suitable for onsite water quality evaluation (Khadr and Elshemy 2017).

Figure 10 depicts the deviation of the water quality indices predicted using six ANFIS models from the calculated WQI values. %parameter_name indicates the graph of the model in which that particular parameter was not considered as one of the inputs. The most evident deviations from WQI values are observed for blue, orange, and yellow lines with one or two spikes of red and green lines. This suggests the importance of BOD, pH, and total coliform as input parameters as compared to water temperature and DO. As BOD, pH, and coliform are associated with the River Ganges

locale, utilizing them as three input parameters for WQI prediction via ANFIS may result in good model reproducibility. The less significant inputs could be removed because it contributes less variance to the WQI forecast.

## Impact of hydro-climatic variations

Although BOD, pH, and coliform turn out to be the most sensitive parameters based on the ANFIS model, all parameters are considered for studying the effects of hydro-climatology on water quality. This is due to all parameters being somewhat sensitive toward WQI. Discharge of the Ganges, air temperature, and rainfall are taken as input, and one by one, all water quality parameters are taken as output for modeling the unknown and uncertain relationship among climate change and water quality using ANN and ANFIS. ANN models with six hidden neurons show that water temperature is significantly affected by rainfall, air temperature, and flow rate of the river with an $R^2$ value of 0.47 (Fig. 11). Although the error of ANN-COLIFORM is the lowest (0.0942), and its $R^2$ value is also very low. Similarly, other ANN models have not achieved good accuracy in the training stage. This indicates ANN's inability to efficiently capture nonlinear relationships and map ambiguous data with time variations.

On the other hand, ANFIS emerges successful in providing better clarity on how water quality depends on flow and climate parameters (Fig. 11) with relatively good $R^2$ values. The highest correlation is found for total
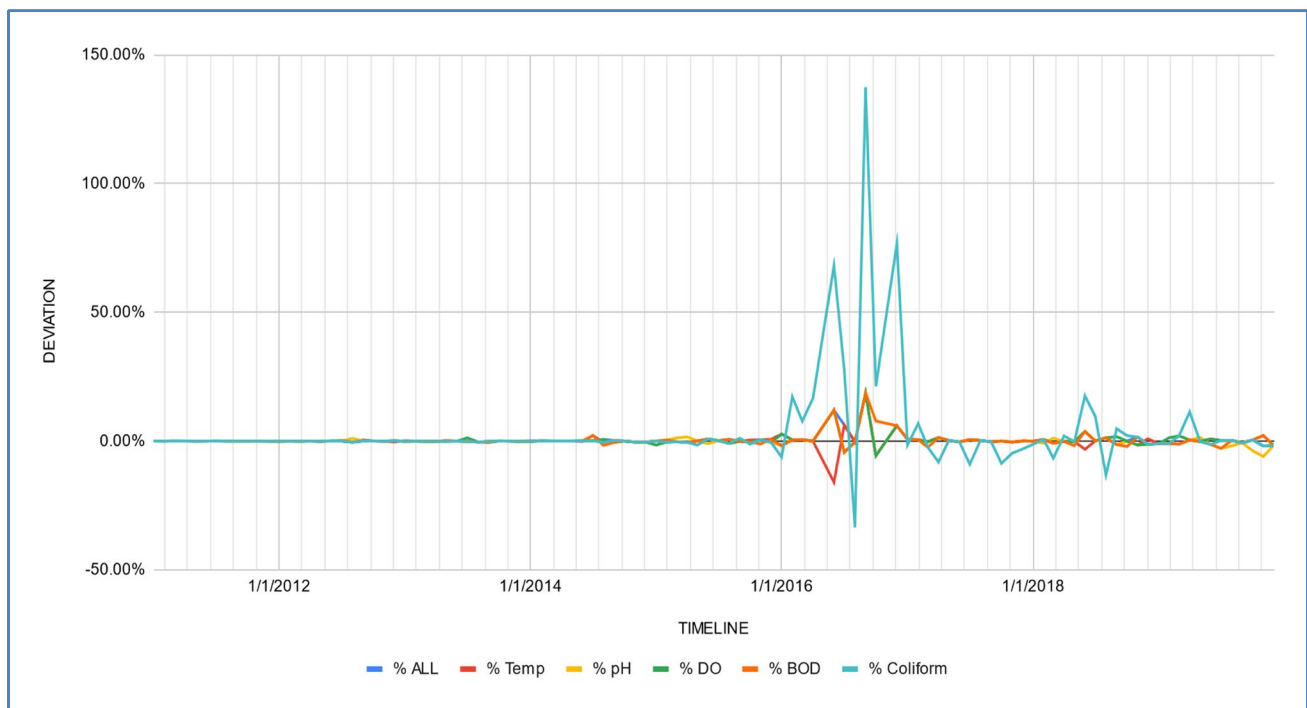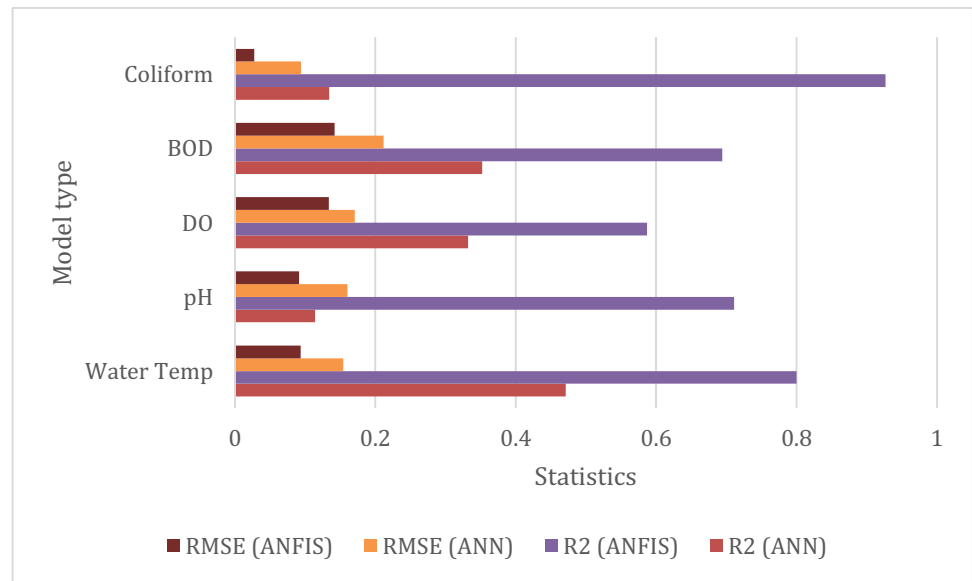


**Fig. 10** Deviation of all ANFIS models in comparison to actual WQI values

**Fig. 11** Comparison of $R^2$ and RMSE values for ANN and ANFIS models for hydro-climatic impact assessment



coliform (0.927) and the least for DO (0.587). Models with DO and BOD as outputs have the least $R^2$ and maximum error values. 'ANFIS-WATER.TEMP' and 'ANFIS-PH' are also comparable based on their RMSE values. Results of ANFIS modeling can be employed for understanding time-dependent uncertain relationships among water quality parameters and climate. In addition, the watershed modelers would also get clarity on framing seasonal policies and planning treatment practices to enhance the water quality under varying climates.

Both the water quality measures and river flow are highly dependent on cumulative antecedent precipitation (Srinivas et al. 2020c). At the Raiwala sampling station, during the rainy season, individual household septic tanks get overcrowded, allowing untreated wastes to spill into drainage ditches and neighboring water bodies, which results in increased coliform concentration. When rain falls on a water source that is not well buffered, runoff can lower the pH of neighboring water. Global warming (increased levels of $CO_2$ in the air) affects the atmospheric temperature and river flows which in turn is responsible for changing pH and DO levels in surface water (Nepal 2016). River water flow has an impact on both low DO and high BOD concentrations. The heavier the rainfall, the more diluted the river water will be, lowering the BOD value (Srinivas et al. 2017). Therefore, studying the impact of climate variations on the water quality of the Ganges is essential.

## Future work

The outcomes of the cloud-based neuro-fuzzy hydro-climatic model proposed in this study have the capability to enhance the existing conventionally available software such as PTMApp (BWSR 2021) and ACPF (ACPF 2021) by

incorporating the uncertainty associated with water quality parameters (Srinivas et al. 2020a). Authors have discussed in detail about how these software would help in developing a decision support system in their previous work (Srinivas et al. 2020b). Some of the prescribed changes in these software are given below.

## PTMApp

- **Individual practices**: PTMApp now has over 20 individual practices based on USDA NRCS practice types (such as farm ponds, saturated buffers, denitrifying bioreactors, and multi-stage ditch) as compared to the 6 treatment groups of practices (source reduction, protection, etc.) in the earlier versions.
- **Economics**: In the original PTMApp, the only upfront cost of installation of the practice was considered. But now, PTMApp outputs that number as well as the total life cycle costs of a practice. Life cycle costs look at the long-term costs for the effective life of the practice and then annualized to get a different perspective on costs. This is helpful when looking at practices like cover crops that have low installation costs but over the long term may cost just as much or more than a structural practice that has high upfront installation costs.

## ACPF

The new version of ACPF (version 4) would be released in Spring 2022. Some of the expected changes include.

- inclusion of economics similar to PTMApp
- addition of pollution reduction benefits and potentially hydrology and carbon benefits as well, and

- also, it would be compatible with the latest versions of ArcGIS.

Looking at the future enhancements of PTMApp and ACPF, there is a great scope to integrate a cloud-based neuro-fuzzy hydro-climatic model to deal with uncertainty and sensitivity associated with costs, locations and sizes of practices, water quality benefits, and stakeholder preferences.

## Conclusions

The study presents various running AI-based models in the cloud to identify sensitive water quality parameters and assess the impact of hydro-climatic conditions on water quality. The collection of data from the monitoring sensors in the river and sending this data to the cloud can enable continuous analysis and assessment. Relationships among water quality and climate change derived using the proposed models would guide the hydro-climatologists to plan advanced treatments based on predicted climatic changes. The results demonstrate that artificial intelligence has been proven effectively equipped for forecasting the future health of the rivers in a real-time network, especially when data is scarce and lacks proper detail. Machine learning models (ANN, ANFIS, ANN-PSO) generate the desired results in a fast, accurate, and inexpensive manner, even with small and imperfect datasets. ANFIS could very well handle the uncertainties associated with water quality parameters. Overall, this study proposes a framework to develop machine learning models (ANN, ANN-PSO, and ANFIS) for predicting the Water Quality Index (WQI) of River Ganges and to assess the impact of climate change on water quality. The novelty of this work lies in evaluating the sensitivity of water parameters with respect to WQI which enhances the model development for WQI prediction. In addition, ANN has been modified using fuzzy logic and PSO algorithms. A unique meta-heuristic method of linking flow rate and climate change parameters with water quality characteristics has been developed using ANN and ANFIS, where ANFIS provided accurate results. ANN, hybrid ANN (weights optimized by advanced PSO), and ANFIS capture the linear relationship between parameters and numerically computed WQI with satisfactory $R^2$ values of 0.999, 0.998, and 0.999, respectively. Sensitivity analysis shows the significance of 'BOD, pH, and total coliform' as inputs for WQI forecasting as compared to water temperature and DO. Moreover, ANFIS is revealed to be a more reliable tool than ANN for assessing the impact of hydro-climatic change on a river's water quality, with a maximum $R^2$ of 0.927 and least NMSE of 0.027 in the case of total coliform. Uncertainty in our study is dealt with by amalgamating fuzzy

inference systems in the ANN model (ANFIS). Results clearly indicate that ANFIS achieves pattern recognition for unknown nonlinear and linear relationships better than ANN and hybrid ANN algorithms. The proposed machine learning framework is flexible and replicable for any number of water quality and climate parameters in any water body in the world. The future scope includes the coupling of classical and probabilistic ways with AI and Geographical Information Systems (GIS) to incorporate uncertainty for large datasets in a more satisfactory manner.

**Author contribution** AJ: methodology, data curation, software, writing – original draft preparation, investigation. RS: conceptualization, data curation, methodology, software, writing – original draft preparation, investigation, writing – reviewing and editing, visualization, supervision. DK: writing – reviewing and editing, conceptualization, visualization. All authors read and approved the final manuscript.

**Data availability** Data could be made available on request.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

ACPF (2021) Agricultural Conservation Planning Framework (ACPF). https://acpf4watersheds.org/about-acpf/. Accessed 25 June 2021.

Aghaarabi E, Aminravan F, Sadiq R, Hoorfar M, Rodríguez MJ, Najjaran H (2017) Application of neuro-fuzzy based expert system in water quality assessment. Int J Syst Assur Eng Manag 8(4):2137–2145

Aghel B, Rezaei A, Mohadesi M (2019) Modeling and prediction of water quality parameters using a hybrid particle swarm optimization–neural fuzzy approach. Int J Environ Sci Technol 16(8):4823–4832

Alizamir M, Sobhanardakani S (2018) An artificial neural network-particle swarm optimization (ANN-PSO) approach to predict heavy metals contamination in groundwater resources. Jundishapur J Health Sci 10(2)

Al-Mukhtar M, Al-Yaseen F (2019) Modeling water quality parameters using data-driven models, a case study Abu-Ziriq Marsh in south of Iraq. Hydrology 6(1):24

Avand M, Moradi H (2020) Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability. J Hydrol 125663

Azad A, Karami H, Farzin S, Mousavi SF, Kisi O (2019) Modeling river water quality parameters using modified adaptive neuro fuzzy inference system. Water Sci Eng 12(1):45–54

Bonyadi MR, Michalewicz Z (2016) Impacts of coefficients on movement patterns in the particle swarm optimization algorithm. IEEE Trans Evol Comput 21(3):378–390

BWSR (2021). Prioritize target and measure application (PTMApp). https://ptmapp.bwsr.state.mn.us/ (Accessed 23 June 2021)

Chatterjee S, Sarkar S, Dey N, Ashour AS, Sen S, Hassanien AE (2017) Application of cuckoo search in water quality prediction using artificial neural network. Int J Comput Intell Stud 6(2–3):229–244

Chen WB, Liu WC (2015) Water quality modeling in reservoirs using multivariate linear regression and two neural network models. Adv Artificial Neural Syst 2015.

CPCB (2013) Pollution assessment: River Ganga. Central Pollution Control Board, New Delhi, India.

CPCB (2016) Restoration/rejuvenation of River Ganga Suggestions/proposals for phase-i, segment 'b' (Haridwar Down to Kanpur Down). Central Pollution Control Board (CPCB), New Delhi, India

CPCB (2019) ENVIS Centre on control of pollution water, air and noise, Central Pollution Control Board (CPCB), New Delhi, India

CWC (2019) Annual report. Ministry of Water Resources, River Development & Ganga Rejuvenation, Central Water Commission (CWC), New Delhi, India.

Dwivedi S, Mishra S, Tripathi RD (2018) Ganga water pollution: a potential health threat to inhabitants of Ganga basin. Environ Int 117:327–338

ENVIS (2016). Status of grossly polluting industries in Uttarakhand. Uttarakhand Environmental Protection and Pollution Control Board.

Fu Z, Cheng J, Yang M, Batista J, Jiang Y (2020) Wastewater discharge quality prediction using stratified sampling and wavelet de-noising ANFIS model. Comput Electr Eng 85:106701

García-Alba J, Bárcena JF, Ugarteburu C, García A (2019) Artificial neural networks as emulators of process-based models to analyse bathing water quality in estuaries. Water Res 150:283–295

Geetha S, Gouthami SJSW (2016) Internet of things enabled real time water quality monitoring system. Smart Water 2(1):1–19

Haritash AK, Gaur S (2016) Assessment of water quality and suitability analysis of River Ganga in Rishikesh, India. Appl Water Sci 6(4):383–392

IMD, 2019. State level climate change trends in India. Indian Meteorological Department, New Delhi, India.

Jha MK, Shekhar A, Jenifer MA (2020) Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index. Water Res 179:115867

Juan C, Genxu W, Tianxu M., Xiangyang S (2017) ANN model-based simulation of the runoff variation in response to climate change on the Qinghai-Tibet plateau, China. Adv Meteorol 2017

Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN (2019) Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. Model Earth Syst Environ 5(3):951–962

Khadr M, Elshemy M (2017) Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt. Ain Shams Eng J 8(4):549–557

Kisi O (2015) Streamflow forecasting and estimation using least square support vector regression and adaptive neuro-fuzzy embedded fuzzy c-means clustering. Water Resour Manage 29(14):5109–5127

Laanaya F, St-Hilaire A, Gloaguen E (2017) Water temperature modelling: comparison between the generalized additive model, logistic, residuals regression and linear regression models. Hydrol Sci J 62(7):1078–1093

Lu H, Ma X (2020) Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere 249:126169

Matta G, Nayak A, Kumar A, Kumar P (2020) Water quality assessment using NSFWQI, OIP and multivariate techniques of Ganga River system, Uttarakhand, India. Appl Water Sci 10(9):1–12

Melesse AM, Ahmad S, McClain ME, Wang X, Lim YH (2011) Suspended sediment load prediction of river systems: an artificial neural network approach. Agric Water Manag 98(5):855–866

Mimikou MA, Baltas E, Varanou E, Pantazis K (2000) Regional impacts of climate change on water resources quantity and quality indicators. J Hydrol 234(1–2):95–109

Mujere N, Moyce, W (2018) Climate change impacts on surface water quality. In Hydrology and Water Resource Management: Breakthroughs in Research and Practice (pp. 97–115). IGI Global.

Nepal S (2016) Impacts of climate change on the hydrological regime of the Koshi River basin in the Himalayan region. J Hydro-Environ Res 10:76–89

Oladipo JO, Akinwumiju AS, Aboyeji OS, Adelodun AA (2021) Comparison between fuzzy logic and water quality index methods: a case of water quality assessment in Ikare community, Southwestern Nigeria. Environ Challenges 3:100038

Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. Swarm Intell 1(1):33–57. https://doi.org/10.1007/s11721-007-0002-0

Pramanik N, Panda RK (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. Hydrol Sci J 54(2):247–260

Rajaee T, Khani S, Ravansalar M (2020) Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. Chemom Intell Lab Syst 200:103978

Sagan V, Peterson KT, Maimaitijiang M, Sidike P, Sloan J, Greeling BA, Maalouf S, Adams C (2020) Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth Sci Rev 205:103187

Şahin M, Erol R (2017) A comparative study of neural networks and ANFIS for forecasting attendance rate of soccer games. Math Comput Appl 22(4):43

Sahu M, Mahapatra SS, Sahu HB, Patel RK (2011) Prediction of water quality index using neuro fuzzy inference system. Water Quality Exposure and Health 3(3–4):175–191. https://doi.org/10.1007/s12403-011-0054-7

Shah KA, Joshi GS (2017) Evaluation of water quality index for River Sabarmati, Gujarat, India. Appl Water Sci 7(3):1349–1358

Shil S, Singh UK, Mehta P (2019) Water quality assessment of a tropical river using water quality index (WQI), multivariate statistical techniques and GIS. Appl Water Sci 9(7)

Sotomayor G, Hampel H, Vázquez RF (2018) Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. Water Res 130:353–362

Srinivas R, Singh AP (2018) An integrated fuzzy-based advanced eutrophication simulation model to develop the best management scenarios for a river basin. Environ Sci Pollut Res 25(9):9012–9039

Srinivas R, Singh AP (2018a) Impact assessment of industrial wastewater discharge in a river basin using interval-valued fuzzy group decision-making and spatial approach. Environ Dev Sustain 20(5):2373–2397

Srinivas R, Drewitz M, Magner J (2020a) Evaluating watershed-based optimized decision support framework for conservation practice placement in Plum Creek Minnesota. J Hydro Elsevier 583:124573

Srinivas R, Singh AP, Dhadse K (2020b) Hydroclimatic river discharge and seasonal trends assessment model using an advanced spatio-temporal model. Stoch Environ Res Risk Assess 34:381–396

Srinivas R, Singh AP, Dhadse K, Garg C (2020c) An evidence based integrated watershed modelling system to assess the impact of

non-point source pollution in the riverine ecosystem. J Clean Prod 246:118963

Srinivas R, Singh AP, Dhadse K, Garg C, Deshmukh A (2018b) Sustainable management of a river basin by integrating an improved fuzzy based hybridized SWOT model and geo-statistical weighted thematic overlay analysis. J Hydrol 563:92–105

Srinivas R, Singh AP, Sharma R (2017) A scenario based impact assessment of trace metals on the ecosystem of river Ganges using multivariate analysis coupled with fuzzy decision-making approach. Water Resour Manage 31(13):4165–4185

Srisaeng P, Baxter GS, Wild G (2015) An adaptive neuro-fuzzy inference system for forecasting Australia's domestic low cost carrier passenger demand. Aviation 19(3):150–163

Stryker J, Wemple B, Bomblies A (2018) Modeling the impacts of changing climatic extremes on streamflow and sediment yield in a northeastern US watershed. J Hydrol: Reg Stud 17:83–94

Sutadian AD, Muttil N, Yilmaz AG,  Perera BJC (2016) Development of river water quality indices—a review. Environ Monit Assess 188(1). https://doi.org/10.1007/s10661-015-5050-0

Tabbussum R, Dar AQ (2021) Performance evaluation of artificial intelligence paradigms—artificial neural networks, fuzzy logic, and adaptive neuro-fuzzy inference system for flood prediction. Environ Sci Pollut Res 28(20):25265–25282

Tiwari S, Babbar R, Kaur G (2018) Performance evaluation of two ANFIS models for predicting water quality Index of River Satluj (India). Adv Civ Eng 2018

Tripathi M, Singal SK (2019) Use of principal component analysis for parameter selection for development of a novel water quality index: a case study of river Ganga India. Ecol Ind 96:430–436

Ucun Ozel H, Gemici BT, Gemici E, Ozel HB, Cetin M, Sevik H (2020) Application of artificial neural networks to predict the heavy metal contamination in the Bartin River. Environ Sci Pollut Res 27(34):42495–42512

UPCB (2020) Sewage treatment plant data. UTTARAKHAND POLLUTION CONTROL BOARD, Government Of Uttarakhand, Dehradun, India.

Wang Y, Chen Y (2014) A comparison of Mamdani and Sugeno fuzzy inference systems for traffic flow prediction. J Comput 9(1):12–21

Yaseen ZM, Ramal MM, Diop L, Jaafar O, Demir V, Kisi O (2018) Hybrid adaptive neuro-fuzzy models for Water Quality Index estimation. Water Resour Manage 32(7):2227–2245

Zanganeh M (2020) Improvement of the ANFIS-based wave predictor models by the particle swarm optimization. J Ocean Eng Sci 5(1):84–99

Zhang C, Lai S, Gao X, Xu L (2015) Potential impacts of climate change on water quality in a shallow reservoir in China. Environ Sci Pollut Res 22(19):14971–14982

Zotou I, Tsihrintzis VA, Gikas GD (2020) Water quality evaluation of a lacustrine water body in the Mediterranean based on different water quality index (WQI) methodologies. J Environ Sci Health, Part A 55(5):537–548