water

MDPI

Article

Exploring Artificial Intelligence Techniques for Groundwater Quality Assessment

Purushottam Agrawal [1], Alok Sinha [1], Satish Kumar [2], Ankit Agarwal [3,4], Ashes Banerjee [5], Vasanta Govind Kumar Villuri [2], Chandra Sekhara Rao Annavarapu [6], Rajesh Dwivedi [7], Vijaya Vardhan Reddy Dera [8], Jitendra Sinha [9] and Srinivas Pasupuleti [10,*]

1 Department of Environmental Science and Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad, Dhanbad 826004, Jharkhand, India; agrawal.2015dr1174@ese.iitism.ac.in (P.A.); alok@iitism.ac.in (A.S.)
2 Department of Mining Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad, Dhanbad 826004, Jharkhand, India; satish.19pr0049@me.iitism.ac.in (S.K.); vgkvilluri@iitism.ac.in (V.G.K.V.)
3 Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India; ankit.agarwal@hy.iitr.ac.in or agarwal@pik-potsdam.de
4 Section Hydrology, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
5 Department of Civil Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India; ashes@iitg.ac.in
6 Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad, Dhanbad 826004, Jharkhand, India; acsrao@iitism.ac.in
7 KIET Group of Institutions, Department of Computer Science and Engineering, Ghaziabad 201206, Delhi-NCR, India; anubhav.dwivedi8@gmail.com
8 SMS India Ltd., Khetri, Rajasthan 333503, India; deravijay@gmail.com
9 Soil and Water Engineering, SVCAETRS, Indira Gandhi Krishi Vishwavidyalaya, Raipur 492012, Chhattisgarh, India; jsvenusmars@gmail.com
10 Department of Civil Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad, Dhanbad 826004, Jharkhand, India
* Correspondence: srinivas@iitism.ac.in

Citation: Agrawal, P.; Sinha, A.; Kumar, S.; Agarwal, A.; Banerjee, A.; Villuri, V.G.K.; Annavarapu, C.S.R.; Dwivedi, R.; Dera, V.V.R.; Sinha, J.; et al. Exploring Artificial Intelligence Techniques for Groundwater Quality Assessment. *Water* **2021**, *13*, 1172. https://doi.org/10.3390/w13091172

Academic Editor: Pankaj Kumar

Received: 8 January 2021
Accepted: 20 April 2021
Published: 23 April 2021

Abstract: Freshwater quality and quantity are some of the fundamental requirements for sustaining human life and civilization. The Water Quality Index is the most extensively used parameter for determining water quality worldwide. However, the traditional approach for the calculation of the WQI is often complex and time consuming since it requires handling large data sets and involves the calculation of several subindices. We investigated the performance of artificial intelligence techniques, including particle swarm optimization (PSO), a naive Bayes classifier (NBC), and a support vector machine (SVM), for predicting the water quality index. We used an SVM and NBC for prediction, in conjunction with PSO for optimization. To validate the obtained results, groundwater water quality parameters and their corresponding water quality indices were found for water collected from the Pindrawan tank area in Chhattisgarh, India. Our results show that PSO–NBC provided a 92.8% prediction accuracy of the WQI indices, whereas the PSO–SVM accuracy was 77.60%. The study's outcomes further suggest that ensemble machine learning (ML) algorithms can be used to estimate and predict the Water Quality Index with significant accuracy. Thus, the proposed framework can be directly used for the prediction of the WQI using the measured field parameters while saving significant time and effort.

Keywords: WQI; Pindrawan tank area; drinking water quality; artificial intelligence; particle swarm optimization; support vector machine; naive Bayes classifier

1. Introduction

A high enough quantity and appropriate quality of freshwater are some of the fundamental requirements for sustaining human life and civilization. Indeed, the tremendous population growth and miraculous achievements in science and technology have increased groundwater utilization for domestic, industrial, and irrigation purposes multiple folds throughout the world over the last few decades. Rapid urbanization, overexploitation, and unscientific waste disposal have also influenced the accessibility and quality of groundwater. Excessive population growth and rapid urbanization have forced the use of chemicals and pesticides for agricultural purposes, which often results in leaching and mixing into the groundwater. As indicated by the World Health Organization (WHO), inappropriate or polluted water causes around 80% of all diseases in human beings. Furthermore, contaminated groundwater quality cannot be improved or re-established by preventing contamination from the source. Therefore, understanding and determining water quality is imperative in the study of water resources and environmental engineering.

Water quality essentially determines the usability of water from a source in terms of the nature and concentration of the impurities present in the sample [1]. As a combined effect of the continuous deterioration in water quality and quantity, approximately one billion people worldwide face a shortage of adequate and safe water supply. These statistics' increasing nature makes it essential to monitor water quality for its efficient management and supply [2,3].

The most efficient method for classifying water quality is using the Water Quality Index (WQI). Water quality is often estimated based on water quality indices [4,5]. It is a tool that has been extensively utilized to assess the performance of water quality management approaches [6]. The approach and methodology used for calculating and interpreting water quality indices have evolved over the years [7–11]. The estimated values of water quality indices have been used to indicate water samples' suitability for day-to-day use. They can be utilized effectively in the execution of water quality overhauling programs.

The WQI's variables comprise biological oxygen demand (BOD), temperature, dissolved oxygen (DO), total suspended solids (TSSs), ammoniacal nitrogen (AN), chemical oxygen demand (COD), and pH [12]. Groundwater quality indices (GQIs) are usually forecasted by measuring the standard variables, such as magnesium (Mg^{2+}), calcium (Ca^{2+}), and nitrate (NO^{-3}) [13–15]. The value provided by the WQI is significant enough to help decision makers. However, estimating the WQI is not that simple because subindex calculations are done in the WQI equations themselves. Several methods are available in the literature for the computation of the WQI worldwide, e.g., United States National Sanitation Foundation Water Quality Index (NSFWQI), the British Columbia Water Quality Index (BCWQI), and the Canadian Water Quality Index (CWQI).

The WQI aims to convert the complicated water quality information into straightforward data that is readily useable by researchers and conveyable to people in general. The calculation process in the case of some approaches applied in several countries, including India [6,16], can be exceptionally intricate and time consuming. As a result, the process always contains the risk of attracting unintended miscalculations [17]. Thus, the limitations for the calculations of WQI are the following: (a) time consuming, (b) lengthy process, (c) complicated process, and (d) different equations are used for WQI calculations, hence there are inconsistencies. It may be obvious from the above discussion that no standard method is available for the WQI.

To conquer the above problems, a few scientists have proposed a nonphysical approach that can successfully predict WQI using machine learning (ML) and artificial intelligence (AI) [18–20]. After satisfactory training, an AI-based model can promptly produce a WQI value by eliminating the sub-index calculations. Awareness of AI algorithms is increasing due to benefits that include nonlinear structures, the capability to calculate complicated trends, the capability to manage huge datasets consisting of different data

scales, and insensitivity to absent data. The forecasting capability of ML–AI algorithms greatly relies on the procedures and exactness of the data collection and analysis. The continuous evolution of computational ability has allowed researchers to use diverse arrangements of ML–AI models. Approaches such as artificial neural networks (ANNs) [17,21–26] adaptive neuro-fuzzy inference systems [27–31], and support vector machines (SVMs) [32] have been effectively applied to predict the quality of water worldwide. Abba et al. (2020) [33] describe in detail the ML–AI techniques that are used for WQI measurement. Most of these ML–AI algorithms can perform with a certain degree of accuracy and it is challenging to compare them based on their performance [25,34].

The AI techniques used in the present study, sometimes include complex manual implementation to reduce its actual effectiveness for water quality management personnel. Practitioners have a great interest in learning the codes such that the codes can be used for solving complex models like the one discussed above. A comprehensive comparison of such models' applications with required software packages must be carried out to improve the accuracy of predictions and the suitability of the AI-based models. However, various data mining programs do not involve vast manipulation of several AI models; instead, the majority of them just support fundamental methods without optimization.

Our study also aimed to develop a user-friendly interface in MATLAB for practitioners that do not have a programming background. The recommended interface is based on a nature-inspired metaheuristic classification system that integrates particle swarm optimization (PSO), along with an SVM and NBC. The water quality was forecasted using fundamental AI techniques, which involved a particle swarm optimization (PSO) algorithm combined with support vector machines (SVMs) for prediction. The classification and predictive AI system investigated in the study was developed using four AI models (single), hybrid metaheuristic regression, and four ensembles (i.e., stacking, voting, bagging, and tiering). The baseline models encompassed single models by using two AI techniques: SVM and NBC, respectively. Subsequently, the ensemble models integrated the registered single models and utilized voting, bagging, tiering, and stacking methods. The goal of the present work was to propose a framework for flexible water quality modeling. The analytical technique had similar goals: the models' predictive accuracy and applicability. The framework will empower administrators and hydrologists to choose the best analytical tools for water management using AI techniques.

These models should be selected based on specific requirements. However, sometimes applying an ensemble model can significantly enhance the model accuracy and reduce the computational cost. In the present study, the combination of the PSO algorithm's applicability with an SVM and NBC was exploited. A framework was proposed for predicting the WQI in the Pindrawan tank area, Raipur region, Chhattisgarh, India.

2. Study Area

The Pindrawan tank command area was the area under study (Figure 1); it is situated within 81°45′–81°50′ E and 21°20′–21°25′ N in the upper Mahanadi River valley (southeastern part) and Raipur district of Chhattisgarh, India. A total of nine villages, namely, Pauni, Amlitalab, Khauna, Deogaon, Bangoli, Dhansuli, Kurra, Baraonda, and Nilja, come under the study area, which has a tropical wet and dry climate. The temperature in this part of India remains moderate throughout the year. The highest temperatures in the year are observed from March to June.

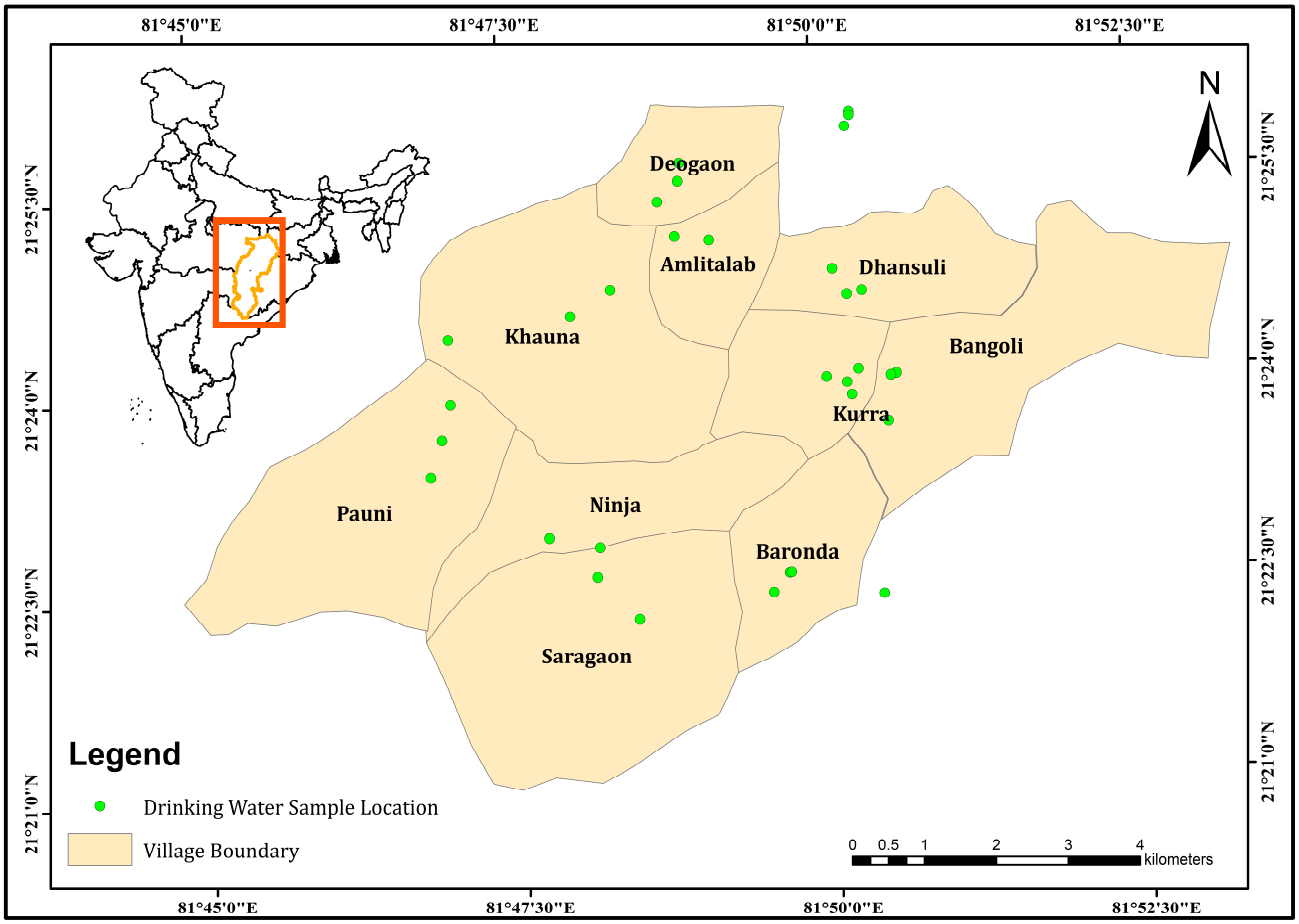


Figure 1. Map of the study area showing the Pindrawan tank command area's geographical location in Chhattisgarh State, India. The figure shows the location of the study area at the country and state levels, as well as the village boundaries that are under the Pindrawan tank command area with drinking water sample locations (green color points).

3. Methodology

3.1. Data Collection and Water Quality Estimation

The groundwater samples were collected in 2018 during the pre-monsoon period from hand pumps and bore wells (37 sites), which are extensively utilized for drinking in the Pindrawan tank area. The identification of the sampling points was performed using topographic sheets and GPS, and the maps were prepared using ArcGIS 10.1 (ESRI, California USA). Topographic sheets were utilized to prepare the base map and recognize the general features of the area. GPS techniques were used to identify the geographic position of each sampling point. The collected groundwater samples were investigated for the concentration of different parameters, namely, electrical conductivity (EC), pH, total dissolved solids (TDSs), total hardness (TH), alkalinity, bicarbonate (HCO_3^-), chloride (Cl^-), sulfate (SO_4^{2-}), nitrate (NO_3^-), fluoride (F^-), calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+), iron (F^-), and chromium (Cr^{2+}), per the specification of the Federation and American Public Health Association (2005). The EC and pH of the collected samples were measured using an EC and pH meter on the field. Fluoride concentrations were analyzed based on the selective electrode method. TH, chloride, and alkalinity were measured using titrimetric methods. Heavy metals were measured using an atomic absorption spectrum and prescribed safety measures were considered to avoid contamination.

The locations of the sampling stations are presented in Figure 1. The concentrations of the parameters were compared with the acceptable limits prescribed by BIS (2012) [35]. The permissible limits of potassium, bicarbonate, and sodium are reported in [36,37].

The WQI of the collected samples was calculated using the weighted arithmetic Water Quality Index (WQI) method [38–40]. The weights (W_i) that were assigned to each parameter according to their impact on the water quality are shown in Table 1.

Table 1. Water quality parameters used when calculating the WQI.

Parameters	Indian Standards	Weight (W_i)	Unit Weight (W_i)	Parameters	Indian Standards	Weight (W_i)	Unit Weight (W_i)
EC	300	1	0.024	Alkalinity	200	3	0.073
PH	6.5–8.5	2	0.049	TH	300	2	0.049
TDS	500	3	0.073	Fluoride	1	4	0.098
Calcium	75	2	0.049	Iron	0.3	4	0.098
Magnesium	30	2	0.049	Chromium	0.05	4	0.073
Potassium	12	2	0.049	Chloride	250	2	0.049
Sodium	200	1	0.022	Bicarbonate	250	3	0.073
Sulfate	200	3	0.073	Total		41	1
Nitrate	45	3	0.073				

Based on the corresponding WQI values, the quality of the groundwater for drinking purposes can be classified into five categories, as presented in Table 2.

Table 2. WQI classification based on the same WQI used by Ramakrishnaiah et al., 2009 [41].

WQI	Class
0–50	Excellent water quality
50–100	Good water quality
100–200	Poor water quality
200–300	Very poor water quality
>300	Unfit for drinking

3.2. Utilization of AI for the Prediction of the WQI

The present study utilized two powerful machine learning approaches for the estimation of the WQI classes by considering the parameter (variables) values as inputs. All 16 variables resembled a variable vector. The analysis was carried out using 1250 variable vectors (250 for each class), which were generated using PSO to contain the whole array of every class. Calibration was conducted using 1250 variable vectors (250 from each class) by applying tenfold cross-validation, and the assessment was done using 250 variable vectors (50 from every class).

3.2.1. Classification and Prediction Using a PSO–SVM Approach Based on the Water Quality Index

The PSO approach is an extremely powerful algorithm that can optimize different model parameters depending on a population's behavior. The approach was proposed by Eberhart and Kennedy in 1995 [42]. The PSO approach has been efficiently used to solve a multitude of nonlinear problems in diversified fields, such as geology [43,44] , landslide analysis [45,46], forest fire mapping [47], and flood modeling [48,49]. The algorithm is initialized with a population of arbitrarily selected solutions between the maximum and minimum range of the parameters. Several advantages of the PSO approach, including the ease of implementation and convergence, fewer parameters, and the use of parallel computing, makes this approach a more comfortable choice compared to other available optimization techniques. The algorithm was developed based on the conduct of a group of fish or birds selecting the smallest path to a food source [50]. The algorithm can improve the exchange of information between samples in a population through an interactive learning process that helps the population arrive at a consistent solution. Each solution is considered as "bird", also known as "particle", in the solution space. Such interactions between members of the population allow this algorithm to demonstrate a robust search proficiency and advanced adaptability to various problems. In PSO, particles (solutions)

will be collected randomly, and then the best particles will be found by renewing the generation. In each generation, each particle is modified using the next two "best" parameters. The first is the best value based on fitness that has been obtained by it until now (fitness parameters are also stored). This value is called individual best value (pbest). Pbest is the best value of thepartile among all the values obtained so far. The other "best" parameter, which comes from the particle swarm analyzer, is the best value that ha been obtained by any particle in the current population. This highest value is called global best (gbest). The movement of the particles is controlled by these optimal values of pbest and gbest. After finding an improved position, they will continue to control the movement of the flock. In the solution space [51], a particle is comprised primarily of two vectors, namely, velocity (Vi) and position (Xi) [52], by using Equations (1) and (2) respectively. Figure 2 gives the PSO algorithm that is used for the particle optimization. The optimization of these two vectors in the dth dimension is performed through the following equations:

$$v_{id}^{t+1} = wv_{id}^{t} + c_1 r_{1d}(pbest_{id}^{t} - x_{id}^{t}) + c_2 r_{2d}(gbest_{id}^{t} - x_{id}^{t}) \quad (2)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (2)$$

where, w is known as the inertia weight. The value of these parameters specifies the number of particles following the current velocity. The parameters c_1 (cognitive coefficient) and c_2 (social coefficient) are known as the acceleration factors. The parameters c_1 and c_2 represent the self-reasoning capability and the ability to acquire information from any particle's contemporary global optimal solution, respectively. r_1 and r_2 are two independent arbitrary parameter numbers in the range [0, 1] [53]. $pbest_{id}^{t}$ and $gbest_{id}^{t}$ are known as the local optimum (best-known position value of any particle i) and the global optimum (optimal value obtained by the swarm of all particles).

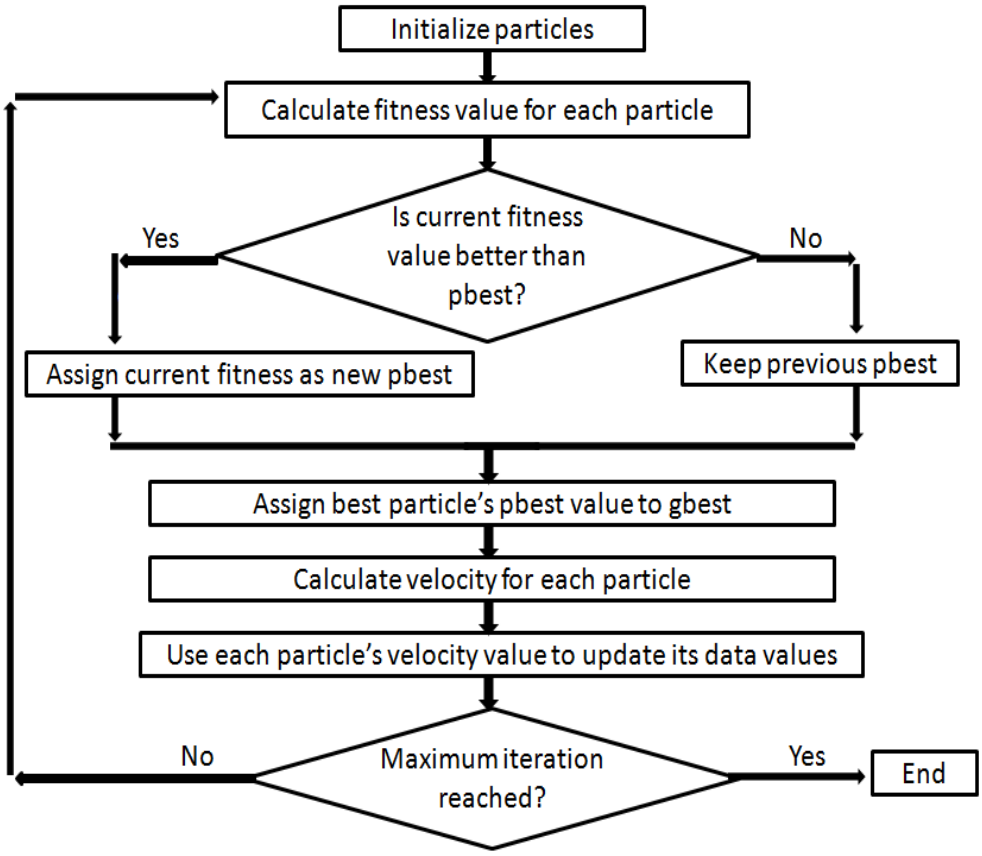


Figure 2. Flowchart for the optimization of the particles.

The coordinate attained by every individual particle in the solution space is recorded by the algorithm. These coordinates are representations of the best solution (fitness value) that has been attained by the particle and is called the local optimum (pbest), whereas the best solution attained by any particle in the vicinity of a specific particle is known as the global optimum (gbest). Although, the particles in the PSO approach tend to move arbitrarily, the best achieved position of the particles (pbest) and the group's best position (gbest) have significant influence over their movement.

Presently, the PSO approach was utilized to produce the optimized values of the WQI, along with all of the 16 water quality variables by considering the variables' lower and upper limits, as presented in Table 3. Based on the corresponding WQI values, the groundwater quality for drinking purposes was classified into five categories (Table 2). To achieve the optimized values of the WQI and water quality variables corresponding to the different classes of water quality, the WQI parameter was considered as the fitness function. The algorithm was set up with an initial population of 50 and processed up to a maximum generation of 500; therefore, a total of 50 × 500 = 25,000 optimized values were generated. The ranges of values for each variable used in the WQI function are presented in Table 3.

Table 3. Comparison of chemical parameters with prescribed standards.

Parameter	Experimentally Obtained Range of Concentration in the Collected Samples	Permissible Limits	Percentage of Samples Exceeding Permissible Limits	Undesirable Effect
pH	7.26–8.59	6.5 to 8.5	2.70	Irritation in eyes, skin, and mucous membranes; skin disorders
EC	152–1998	300	89.19	Cardiac dysrhythmias
TDS (mg/L)	98.8–1199	500	21.62	Gastrointestinal irritation
Alkalinity (mg/L)	60–335	200	29.73	Unpleasant and harmful to aquatic life and humans
Chloride (mg/L)	20–330	250	8.11	Salty taste
Calcium (mg/L)	4–60.5	75	0	Scale formation
Magnesium (mg/L)	4–20.2	30	0	Cerebrovascular disease (Yang, 1998)
Potassium (mg/L)	0–30.9	12	16.20	Bitter taste
Sodium(mg/L)	1.2–18.3	200	0	High blood pressure
Nitrate (mg/L)	3.4–8.2	45	0	Methemoglobinemia
Sulfate (mg/L)	25–50	200	0	Laxative effect
Bicarbonate (mg/L)	2.5–6.5	250	0	Vomiting, dehydration, chronic obstructive pulmonary disease
Fluoride (mg/L)	0.25–0.84	1	0	Mottling of teeth, deformation of bones
Iron (mg/L)	0.015–0.785	0.3	5.41	Diabetes, hemochromatosis, stomach problems, nausea, and vomiting
Chromium (mg/L)	0.007–0.737	0.05	56.76	Hearing loss, blood disorders, hypertension, death at high levels
TH (as mg/L)	138–320	200	43.24	Scale formation in pipes anencephaly, urolithiasis, parental mortality

The procedure for generating the optimal variables' values was as follows:

Step 1—The fitness function was explained using the WQI function, initializing "50 as population" and "500 as the maximum generation."

Step 2—Each variable's maximum and minimum limits were set while using the WQI function according to Table 3.

Step 3—Every particle's movements were recorded in every generation in the vector form comprising the value of the WQI, together with the subsequent values of the 16 variables.

Step 4—The category (class) of each variable vector was obtained by considering its corresponding WQI, as presented in Table 2.

Step 5—A total of 250 variables vectors were selected from each category in such a manner that the entire range of the particular category should be covered, as given in Table 2.

In every generation, the populace shifted from the initial position to a new appropriate place and produced new fitness values. Every particle's movement in every generation was recorded in the vector form containing the WQI value along with the subsequent variables' values. Every random particle updated its fitness value (WQI) in each generation, which was stored in the database and related variables. In PSO, the population's values (swarm) and max iteration (generation) depend on the user. The flowchart for this work is shown in Figure 3. The classification of the WQI values was performed using a support vector machine and a naive Bayes classifier. Before proceeding with the classification, the dataset was normalized between 0 and 1 to enhance the accuracy. The variables' values in vector format were treated as a feature vector in the normalized dataset.

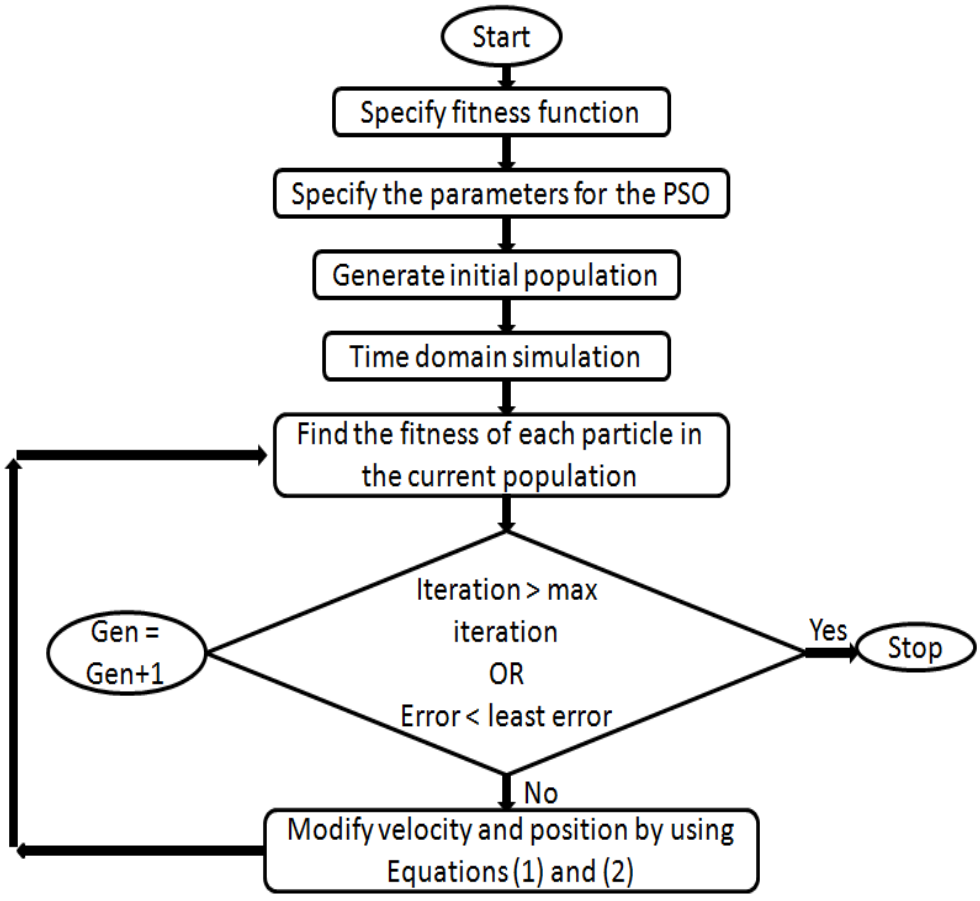


Figure 3. Flowchart describing the workings of the PSO.

3.2.2. Classification Using a Support Vector Machine

The SVM classifier[54] plays an essential and comprehensive role in classification due to its high accuracy and ability to deal with high-dimensional data. The simple form of the classification is the binary used for separating two types of objects belonging to positive (+1) and negative (−1) classes. A support vector machine uses two kinds of concepts

to distinguish between two classes: (1) separation from the margin and (2) the kernel function.

The simple two-dimensional data can be classified by using a straight line. The points that fall above the line belong to one class, and the points that fall below the line belong to another class. The high-dimensional data can be classified by using the hyperplanes. However, in a binary classification, multiple planes can be drawn such that they separate the data into two classes. As such, which plane will be selected for the classification? In this case, the hyperplane that gives the maximum margin will be selected for classification. Therefore, we choose the hyperplane such that the distance from it to the nearest data point on each side is maximized. The classification of the data with the best margin hyperplane is shown in Figure 4.

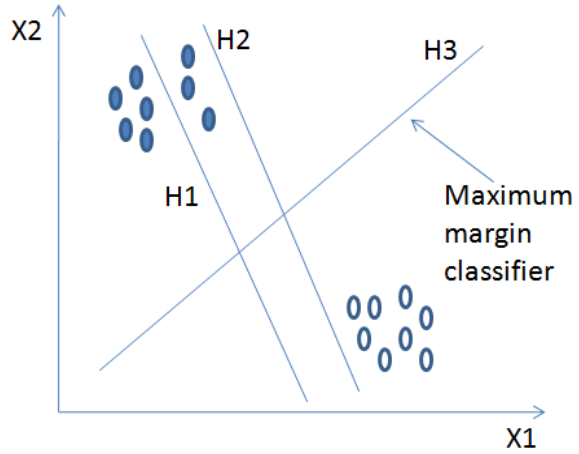


Figure 4. Classifications of data using various hyperplanes.

In Figure 4, there are two types of data points: filled and unfilled dots. Three planes exist, which are named H1, H2, and H3. H1 does not successfully classify the data points. Planes H2 and H3 are both capable of classifying data points, but H2 gives a smaller margin than plane H3.

This is why plane H3 is selected for the classification. Sometimes the data is not classified by hyperplanes because of its distribution in a vast space. In that case, we use a nonlinear separation for the classification. The SVM classifier can efficiently perform this nonlinear classification by using kernel functions. The nonlinear classification is presented in Figure 5. In Figure 5, there are two types of objects, as identified by the solid and hollow dots. The objects represented in this figure cannot be separated using a linear hyperplane; the support vector machine performs this task using kernel functions. The kernel function separates the data in the feature space by using a linear hyperplane.

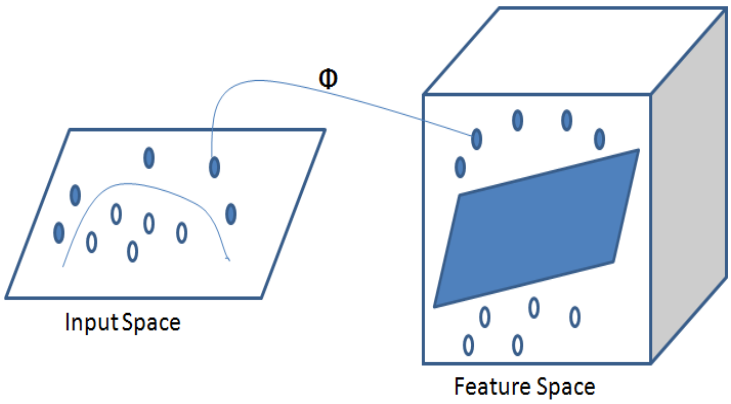


Figure 5. Use of the kernel function in an SVM.

In this work, the SVM classifier separates the individual water quality classes with hyperplanes by using the radial basis kernel (Gaussian) function [55–58]. The distance of a feature vector from the hyperplanes determines its probability of featuring in a specific class. The normalized dataset and the class labels were used as inputs in the present study. The dataset was randomly divided 80:20, where 80% of the dataset was used for training purposes using tenfold cross-validation. In the tenfold cross-validation, the entire dataset was divided randomly into ten equal-sized subsamples. A single subsample was used for testing purposes, and nine subsamples were used for training purposes on ten subsamples. This process was repeated ten times until each of the 10 subsamples were used exactly once for testing purposes. The remaining 20% of the dataset was used for testing and validation purposes.

3.2.3. Classification Using Naive Bayes Classifier

Naive Bayes classifiers are based on Bayes Theorem with a family of algorithms with the same principle, i.e., each pair of features being categorized is independent of every other. The fundamental naive Bayes assumption is that every feature makes an unbiased and identical contribution to the outcome. A naive Bayes classifier is a probabilistic machine learning model that is used for a classification task. The crux of the classifier is based on Bayes' theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad (3)$$

By using Equation (3), the probability of event A happening can be measured by considering that event B has occurred. Here A is the hypothesis and B is the evidence. One assumption that is considered here is that all features are independent/autonomous, which means the presence of one particular feature does not affect the other. Hence it is called naive. Before the PSO–NBC analysis, the dataset was normalized to enhance the performance of the model. A total of 80% of the dataset was used to train the algorithm, whereas 20% of the dataset was used to study the algorithm's prediction accuracy. In this work, continuous values that were associated with each feature were assumed to be distributed according to a Gaussian/normal distribution.

4. Results and Discussions

4.1. Water Quality Index (WQI) Analysis of the Field-Based Samples

The concentration, distribution, and impact of different physicochemical parameters observed from water samples collected from the Pindarwan tank area are discussed in this section. The ranges of concentrations observed for various parameters and the percentages of total samples exceeding the prescribed limit are presented in Table 3, along with their undesirable effect on groundwater quality and human physiology. This section provides an overview of the spatial distribution of the physicochemical parameters that were measured in the Pindarwan tank area; a more detailed description is provided in Figures A1–A15 in the Appendix A.

Out of 37 samples, 32.43% of the samples had excellent water quality, 43.24% of the samples had good water quality, 21.62% of the samples had poor water quality, and 2.71% of the samples had very poor water quality. This may be due to the heavy concentrations of metals, such as Pb and Cr, due to nearby industries, which involve mining activities, thermal power plants, etc. The areas corresponding to these WQI values are presented in Figure 6.

Figure 7 represents a correlation plot between the WQI and the parameters observed from the study area's water samples. The correlation between the independent parameters can be neglected in the plot since these plots are mostly empirically based on specific values. In decreasing order, the influence of different parameters can be presented as chro-

mium, sodium, fluoride, potassium, chloride, conductivity, total dissolved solids, alkalinity, bicarbonate, and pH. Contributions from the rest of the parameters on the overall water quality were much less compared to these parameters. Through observing Figure 7, it can be concluded that water quality for drinking was susceptible to heavy metal concentrations, such as chromium.

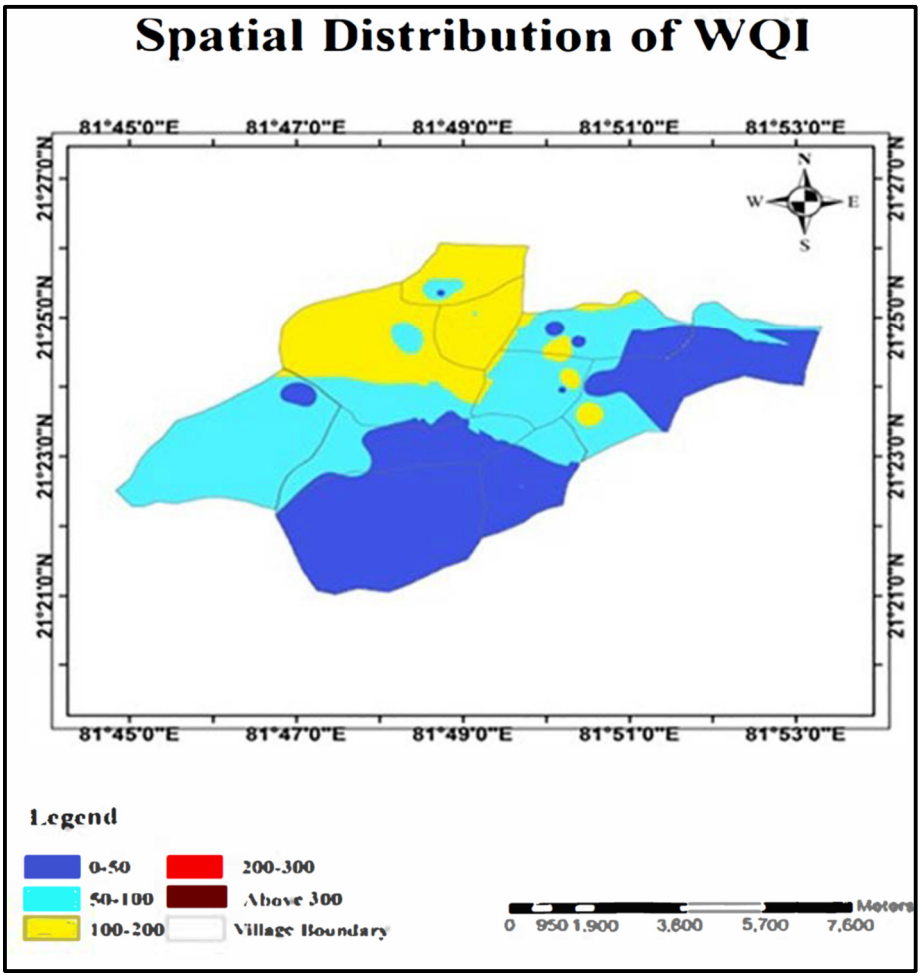


Figure 6. Spatial distribution of the WQI.

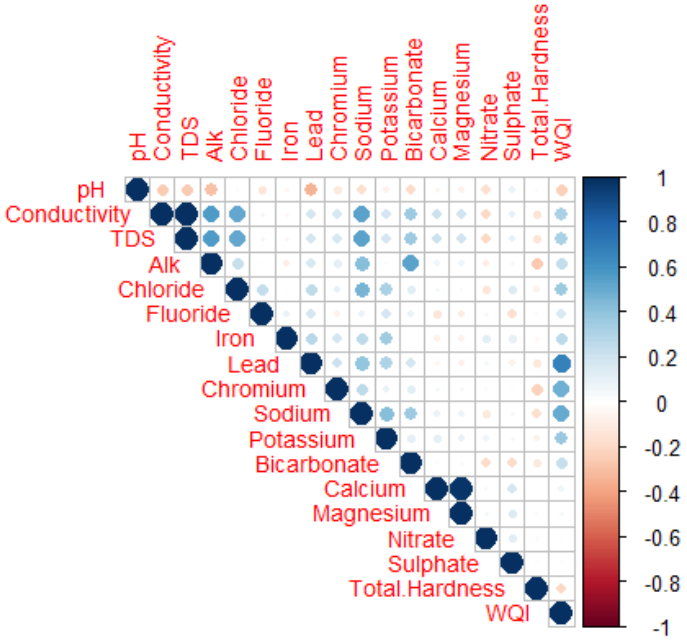
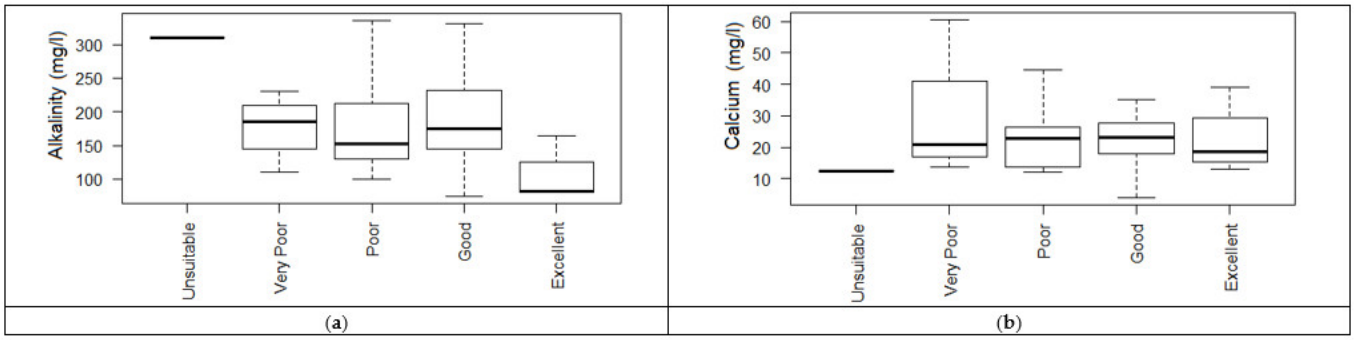


Figure 7. Correlation plot between various groundwater quality parameters.

Based on the WQI, the sample area's drinking water quality was divided into four categories. No sample was observed to be unsuitable for drinking based on the analysis. Very poor water quality was observed from the Raikheda pond area due to a very high chromium concentration. Poor water quality was observed in significant parts of the Deogaon, Dhansuli, Bangoli, Amlitalab, and Khauna villages. Most areas of all the villages had good water quality. Excellent water quality was observed in Saragaon, Nilja, Dhansuli, Bangoli, Khauna, Baronda, and Pauniarea. The observed water qualities may suggest that most of the study area's water quality is satisfactory and there is no immediate danger for the population. However, the values of certain parameters, such as the chromium concentration, total hardness, and total dissolved solids, were alarmingly high for many areas and could become worse. This may significantly influence the present scenario of the water quality in the study area under consideration. Therefore, concerned authorities should note the situation and plan proper steps for maintaining or improving the current situation of the drinking water quality in the study area.

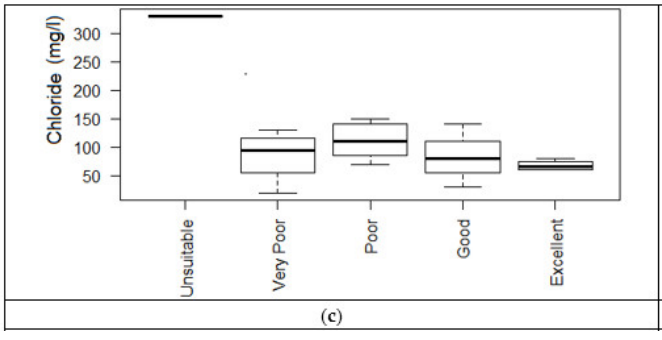
Furthermore, the averages and ranges of the values of different parameters corresponding to water quality are presented in a boxplot format in Figure 8a–p. The concentration of some parameters such as alkalinity, chloride, conductivity, chromium, iron, bicarbonate, sodium, and TDSs are found to be directly proportional and has much more significant impact on the WQI of the study area. These are, therefore, the parameters that have to be first taken care of when aiming to improve the water quality for the specific study area. The influences presented in Figure 8a–p are the combined effect of the concentration of each parameter and the relative weight of each parameter. Therefore, even if a parameter's relative weight is much less, it could make a significant impact if it had a very high concentration. However, these plots are strictly applicable to the present study area and no inference should be derived from these plots for any other samples. The boxplots and correlation plots can be extremely useful for conveying a detailed picture regarding the water quality of the study area and the influence of different parameters on the water quality.


Alkalinity (mg/l)
300
250
200
150
100
Unsuitable
Very Poor
Poor
Good
Excellent

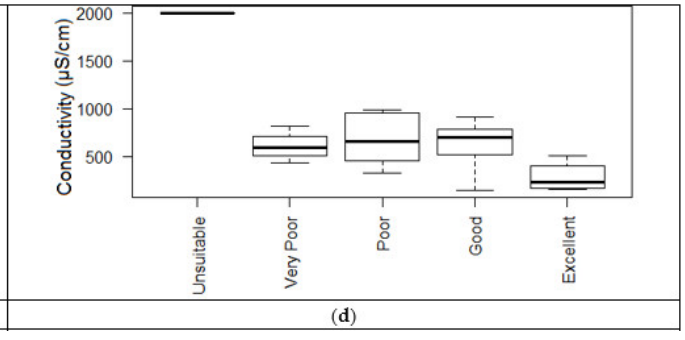
(a)

Calcium (mg/l)
60
50
40
30
20
10
Unsuitable
Very Poor
Poor
Good
Excellent

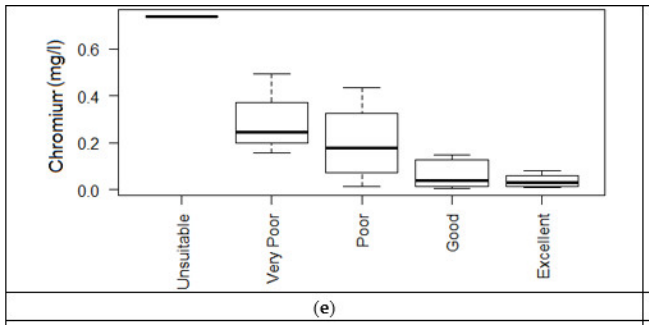
(b)


Chloride (mg/l)
300
250
200
150
100
50
Unsuitable
Very Poor
Poor
Good
Excellent

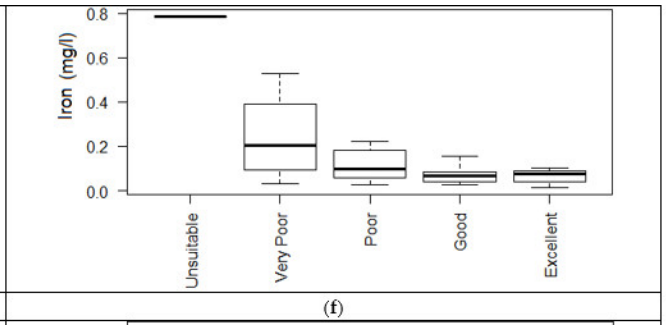
(c)


Conductivity (µS/cm)
2000
1500
1000
500
Unsuitable
Very Poor
Poor
Good
Excellent

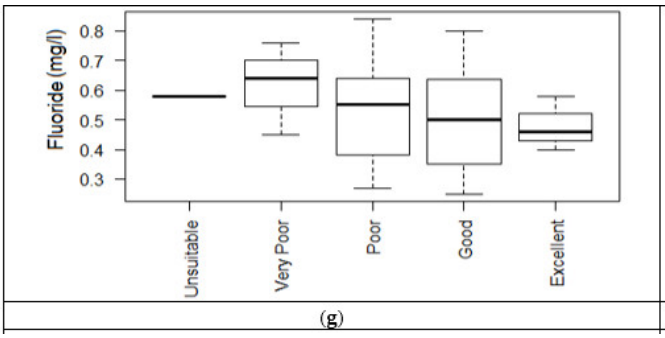
(d)


Chromium (mg/l)
0.6
0.4
0.2
0.0
Unsuitable
Very Poor
Poor
Good
Excellent

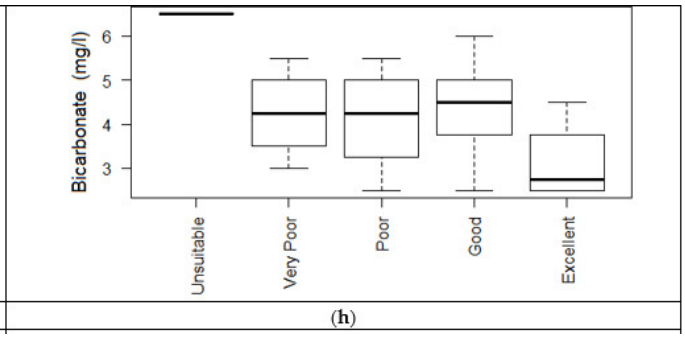
(e)


Iron (mg/l)
0.8
0.6
0.4
0.2
0.0
Unsuitable
Very Poor
Poor
Good
Excellent

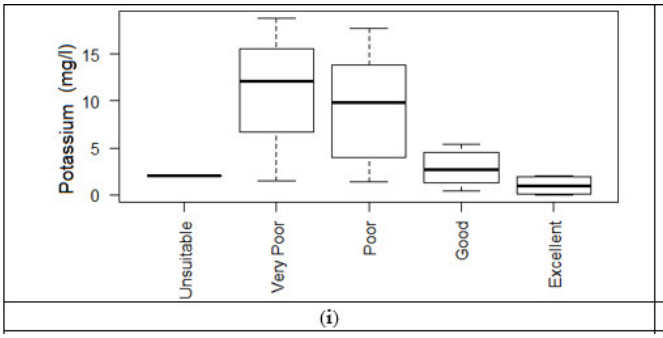
(f)


Fluoride (mg/l)
0.8
0.7
0.6
0.5
0.4
0.3
Unsuitable
Very Poor
Poor
Good
Excellent

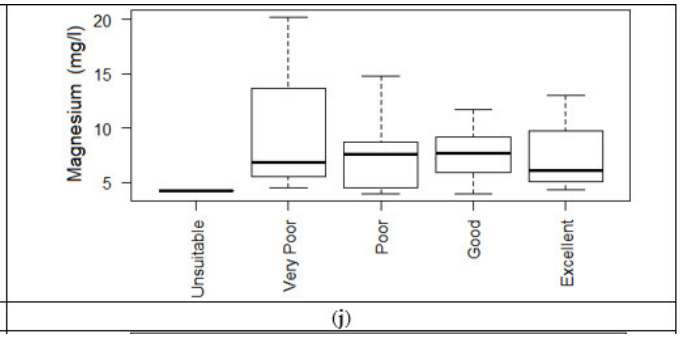
(g)


Bicarbonate (mg/l)
6
5
4
3
Unsuitable
Very Poor
Poor
Good
Excellent

(h)


Potassium (mg/l)
15
10
5
0
Unsuitable
Very Poor
Poor
Good
Excellent

(i)


Magnesium (mg/l)
20
15
10
5
Unsuitable
Very Poor
Poor
Good
Excellent

(j)

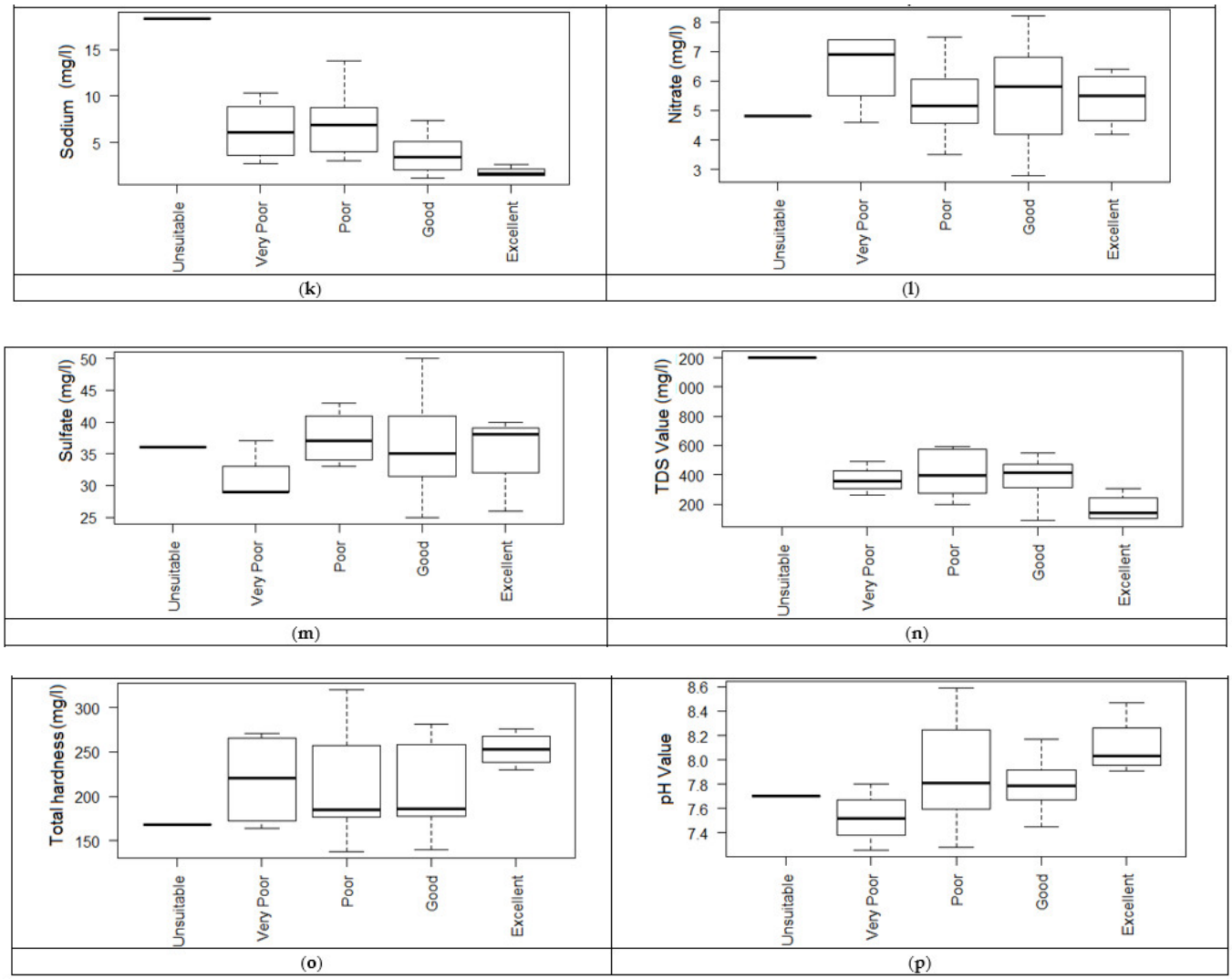


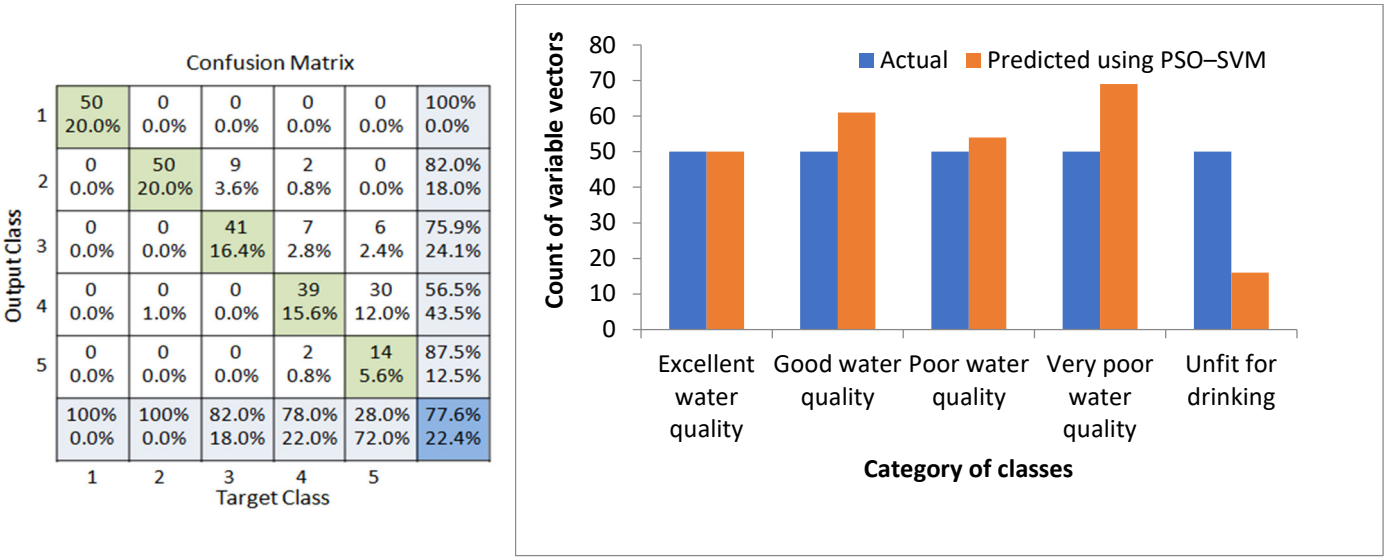
Figure 8. Ranges of various parameters corresponding to the water quality: (**a**) alkalinity, (**b**) calcium, (**c**) chloride, (**d**) conductivity, (**e**) chromium, (**f**) iron, (**g**) fluoride, (**h**) bicarbonate, (**i**) potassium, (**j**) magnesium, (**k**) sodium, (**l**) nitrate, (**m**) sulfate, (**n**) TDSs, (**o**) total hardness, and (**p**) pH.

4.2. Result from the PSO–SVM Study

The performance of the model is presented using the confusion matrix in Figure 9a. The confusion matrix is used to explain the model's classification and overall performance on the testing datasets whose original labels are known. The instances in a predicted class and actual class are represented in every row and each column respectively (or vice versa). In Figure 9a, the rows from the top to the bottom correspond to the excellent, good, poor, very poor, and unfit for drinking water qualities, respectively, as predicted using the SVM classifier.

Furthermore, the columns from left to right follow a similar arrangement of the target class (actual classifications based on the WQI values). Each column related to these classes had 50 variable vectors (water quality class from excellent to unfit for drinking), totaling 250 variable vectors. In the first row, 50 variable vectors are presented, indicating 50 excellent water class WQIs, where the system predicted them all as being in the excellent category. Similarly, in the second, third, fourth, and fifth rows, a sum of 61, 54, 69, and 16 variable vectors are presented, respectively. The result indicates that the algorithm predicted 61 samples as good quality, 54 as poor quality, 69 as very poor quality, and 16 as unfit for the drinking category. The prediction accuracies corresponding to each class are

also presented in the last column from the left-hand side. The overall accuracy of the algorithm was found to be 77.60%. Furthermore, a difference between the classifications based on actual values of the WQI and the predicted classification based on the SVM classifier is presented in Figure 9b.



(**a**) (**b**)

Figure 9. Comparison between the predicted class and target class using the SVM approach: (**a**) confusion matrix and (**b**) column plots.

4.3. Discussion of the PSO–NBC Approach

The PSO–NBC study was carried out by considering the same dataset as in the PSO–SVM approach. The test accuracy is discussed using the confusion matrix presented in Figure 10a. The rows and columns marked as 1 to 5 indicate the excellent (1), good (2), poor (3), very poor (4), and unfit for drinking (5) water qualities. The 51 variable vectors in the first row indicate that the algorithm identified 51 variable vectors as excellent water quality when there were 50 actual excellent water categories (1 more due to misclassification). Similarly, in the second (50 variable vectors of good water quality), third (50 variable vectors of poor water quality), fourth (50 variable vectors of very poor water quality), and fifth rows (50 variable vectors of unfit for drinking water quality), the algorithm placed 57 (good water quality), 46 (poor water quality), 51 (very poor water quality), and 45 (unfit water quality) variable vectors. The prediction accuracy of the algorithm corresponding to each class is presented in the sixth column. The total accuracy of the algorithm was observed to be 92.80%.

The comparisons of the model-predicted outcomes against the actual WQI values are graphically represented in Figure 10b.

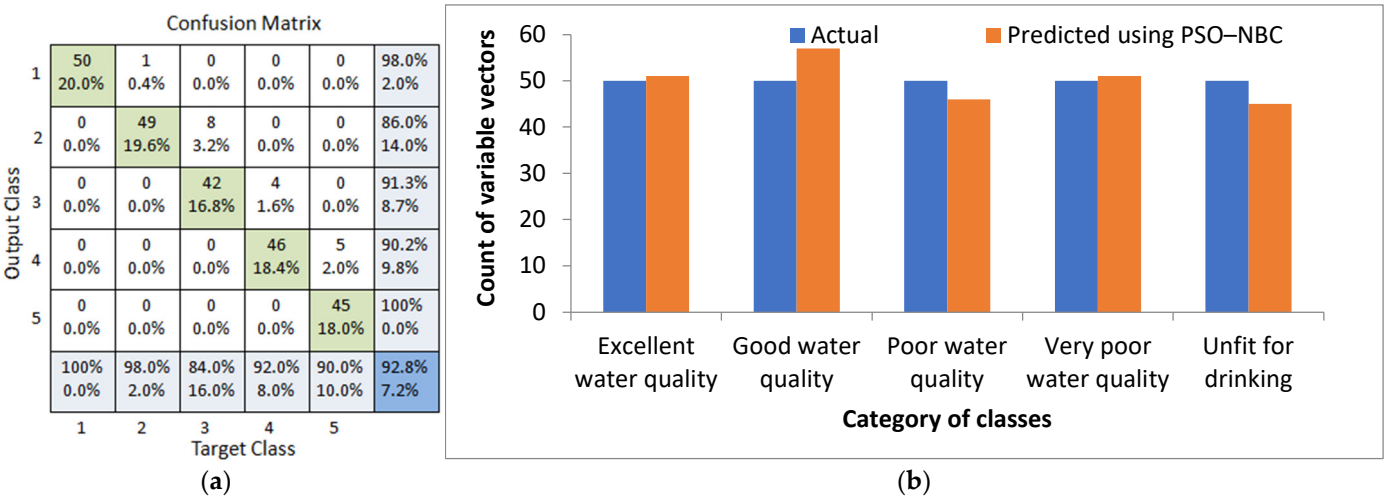


Figure 10. Comparison between the predicted class and target class using the NBC approach: (**a**) confusion matrix and (**b**) column plots.

4.4. Comparison between the PSO–SVM and PSO–NBC Approaches

The performances of the PSO–SVM and PSO–NBC approaches used in the present study are presented in Figure 11.

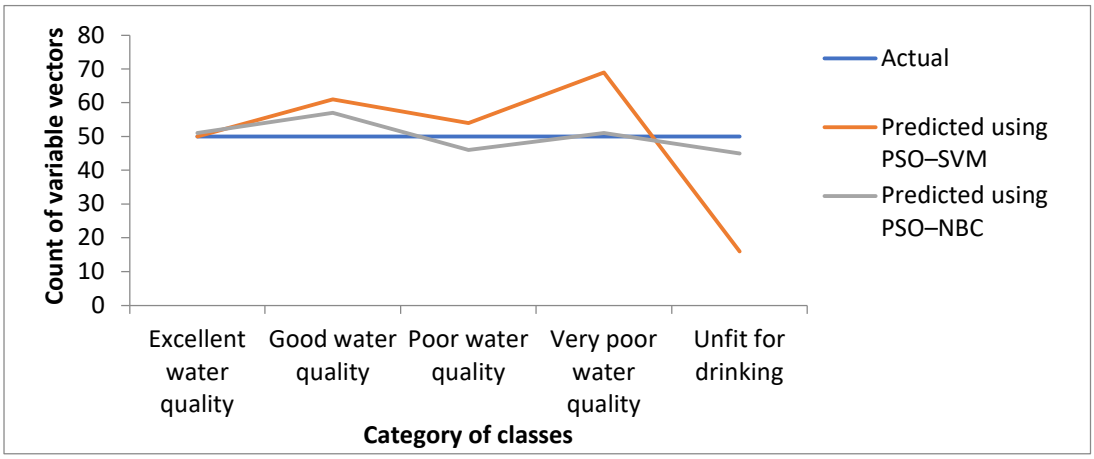


Figure 11. Comparison of the predicted outcomes using the PSO–SVM and PSO–NBC approaches.

The figure indicates that the PSO–SVM algorithm predicted some classes (excellent and poor water categories) with significant accuracy; however, significant deviations were observed in the model's performance for the other categories. On the other hand, the prediction accuracies of PSO-NBC were much higher for all the classes and did not distinctly deviate for any specific categories. Therefore, a naive Bayes classifier aided by particle swarm optimization can be efficiently used to construct a machine learning model to classify water for drinking purposes.

5. Conclusions

The process of WQI estimation is often associated with handling large quantities of identical data. This can create significant confusion during the calculation process and make decision making difficult. A machine-learning-based predictive model can assemble the necessary information and predict the groundwater quality with significant accuracies. This study aimed to utilize modern machine learning techniques for the prediction of water quality for drinking. The groundwater samples collected from parts of the Pindrawan tank command area were used for testing and validation of the developed model.

The collected samples were tested for different parameters of water quality and the subsequent values of WQI were computed. Conclusions derived from the present work are as follows:

1. The calculated WQI values suggested that 32.43% and 43.24% of the water samples of the study area represented excellent and good water qualities, respectively. Similarly, it can also be observed that 21.62% and 2.71% of the water in the study area were of poor and very poor drinking water qualities. Very poor water quality was observed from the Raikheda pond area due to very high chromium concentration. Poor water quality was observed in significant parts of the Deogaon, Dhansuli, Bangoli, Amlitalab, and Khauna villages.
2. The major cation and anion data revealed that all anions were within the limits, except for potassium, where 13% of the samples exceeded the limit. However, the heavy metals pollution in the area due to mining activities could be a cause for concern soon. A total of 48.6% of the samples from the area exceeded the permissible limits of chromium, which can cause conditions such as hearing loss, blood disorders, hypertension, and death at high levels.
3. The study further suggests that ensemble machine learning algorithms can be used for the estimation and prediction of a WQI with significant accuracies. In the present study, a particle swarm optimization approach coupled with a naive Bayes classifier provided a 92.8% accurate prediction of the WQI indices. Therefore, with the help of a user interface, this algorithm can be efficiently utilized for the estimation of WQIs, which can save significant effort and time.

The general outcomes from the present research indicate the benefits of using ensemble machine learning techniques, where outcomes from several different algorithms can be combined and used to achieve predictions with enhanced accuracies. Finally, with the help of a user interface, the algorithm developed in the present study can be used for water quality estimation in different regions across the globe.

The classification in the present study was carried out by taking the synthetic dataset that was generated using particle swarm optimization. However, the developed approach can be further improved if more real data is available. Therefore, the authors suggest using a larger field dataset to obtain better accuracy, though this is often a difficult undertaking provided the painstaking process of sample collection and laboratory analysis for all the water quality parameters. The developed algorithm can be further improved by studying its performance and fine-tuning it with different input parameters.

Author Contributions: Conceptualization, P.A., A.S. and S.P.; data curation, P.A.; formal analysis, P.A. and S.K.; investigation, P.A., S.K. and A.B.; methodology, P.A., A.S., S.K., A.A., A.B. and S.P.; project administration, A.S., V.G.K.V. and S.P.; resources, P.A., A.A., and J.S.; software, P.A., A.B., C.S.R.A. and R.D.; supervision, A.S., J.S., and S.P.; validation, P.A., A.B., C.S.R.A. and R.D.; visualization, A.S., A.A., V.V.R.D., and S.P.; writing—original draft, P.A., A.S., A.B., C.S.R.A. and J.S.; writing—review and editing, A.S., S.K., A.A., V.G.K.V., V.V.R.D., J.S. and S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data may be obtained from the authors upon request.

Acknowledgments: The authors would like to sincerely thank the Indian Institute of Technology (Indian School of Mines) authorities, Dhanbad, for extending their support and allowing the use of facilities from various engineering departments, i.e., Environmental Science and Engineering, Civil Engineering, Mining Engineering, and Computer Science Engineering Departments. The authors would also like to acknowledge the support received and facilities used from the Water Resources Department, Government of Chhattisgarh, and Indira Gandhi Krishi Vishwavidyalaya, Raipur, in

carrying out this research work. A.A. acknowledges the infrastructural support provided by the Indian Institute of Technology Roorkee.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Locations used for the groundwater samples.

Sample No.	Lat	Long	Sample No.	Lat	Long	Sample No.	Lat	Long
1	81.8282	21.3761	13	81.8367	21.3994	25	81.8155	21.4273
2	81.8023	21.3764	14	81.8373	21.3977	26	81.8124	21.4226
3	81.8077	21.371	15	81.834	21.4001	27	81.8152	21.4252
4	81.7961	21.3815	16	81.828	21.3760	28	81.8584	21.4041
5	81.8028	21.3801	17	81.8258	21.3736	29	81.8377	21.4311
6	81.7961	21.3815	18	81.8282	21.3761	30	81.8566	21.4033
7	81.8391	21.4107	19	81.7824	21.3942	31	81.8001	21.4089
8	81.8353	21.4134	20	81.7807	21.3896	32	81.8426	21.4000
9	81.8371	21.4103	21	81.7837	21.3985	33	81.8405	21.3729
10	81.8383	21.4010	22	81.7837	21.4066	34	81.8384	21.4329
11	81.842	21.3943	23	81.8001	21.4089	35	81.8384	21.4325
12	81.8433	21.4002	24	81.8056	21.4119	36	81.819	21.4177
						37	81.8145	21.4183

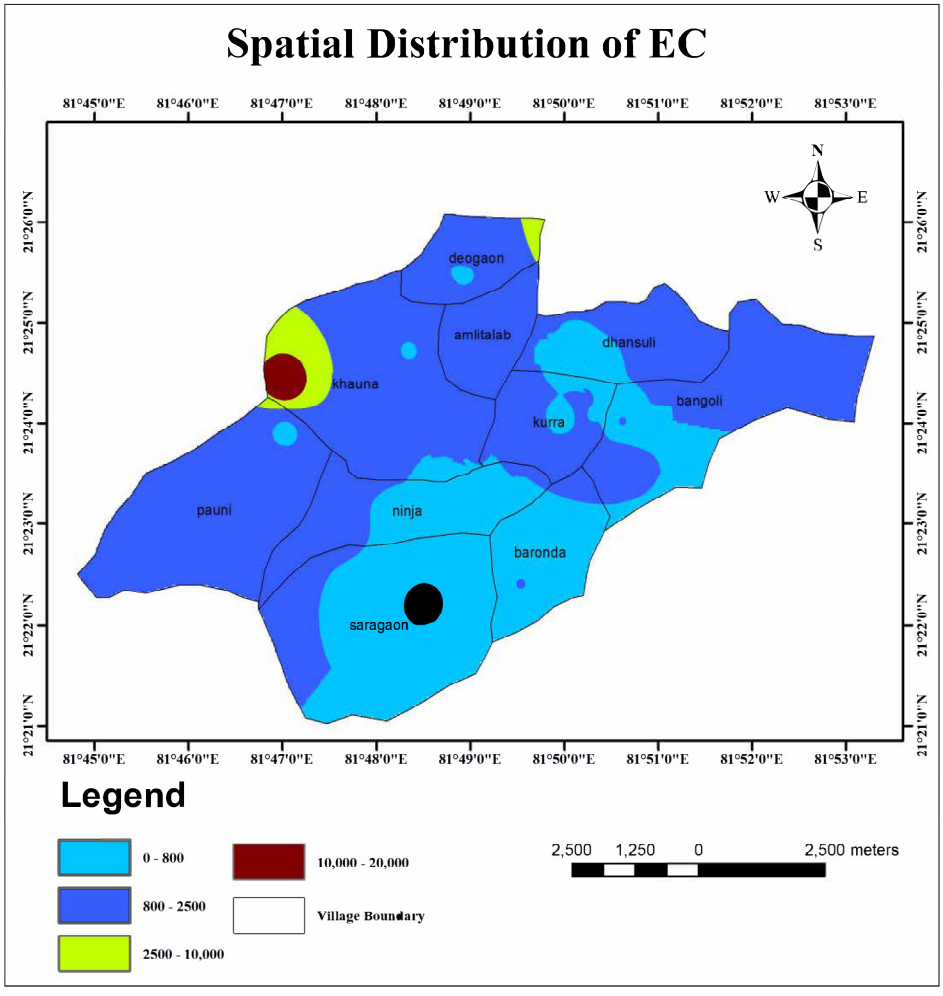


Figure A1. Spatial distribution of EC.

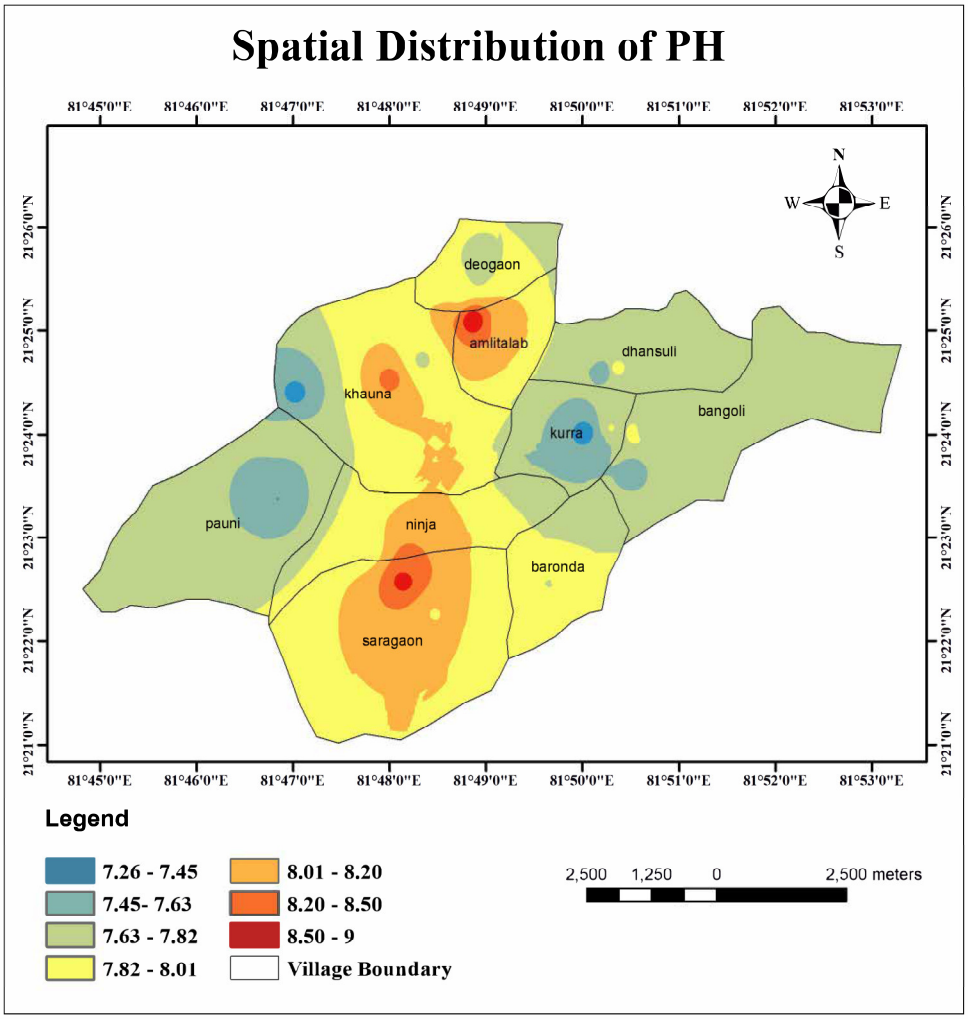


Figure A2. Spatial distribution of PH.

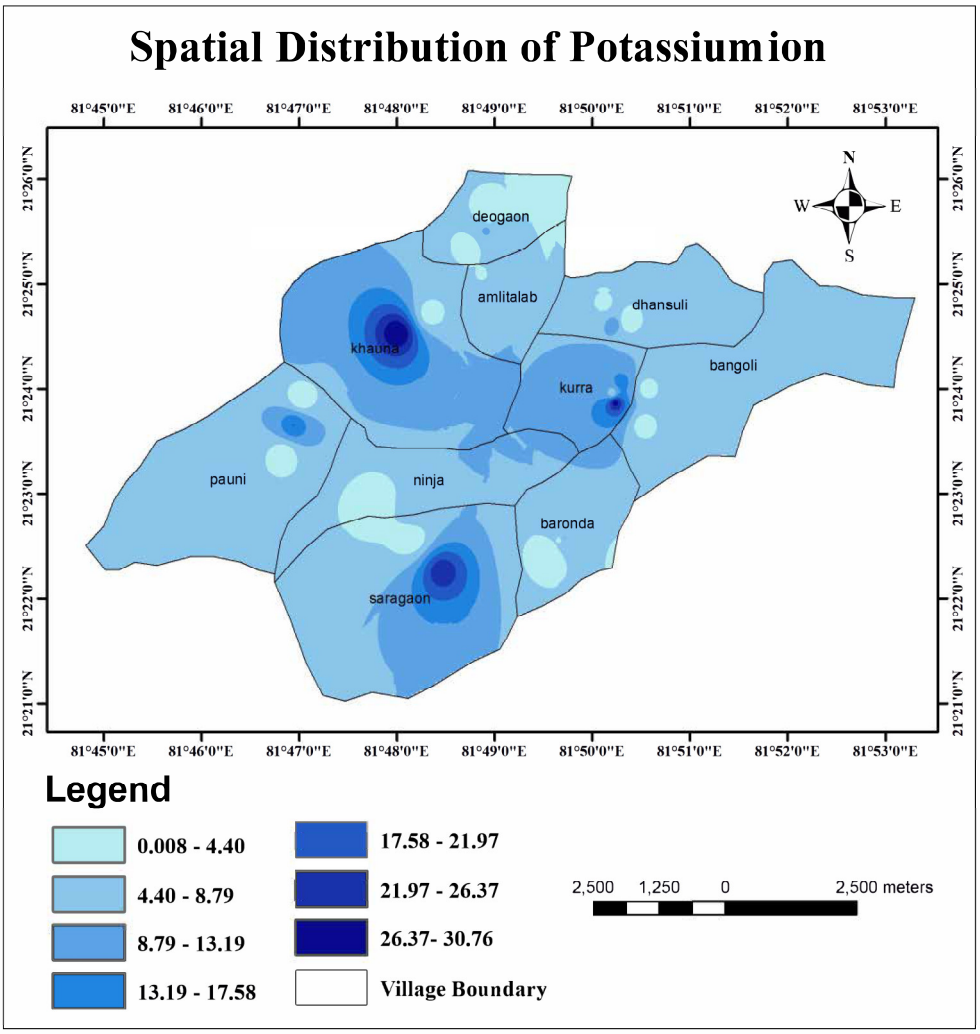


Figure A3. Spatial distribution of potassium.

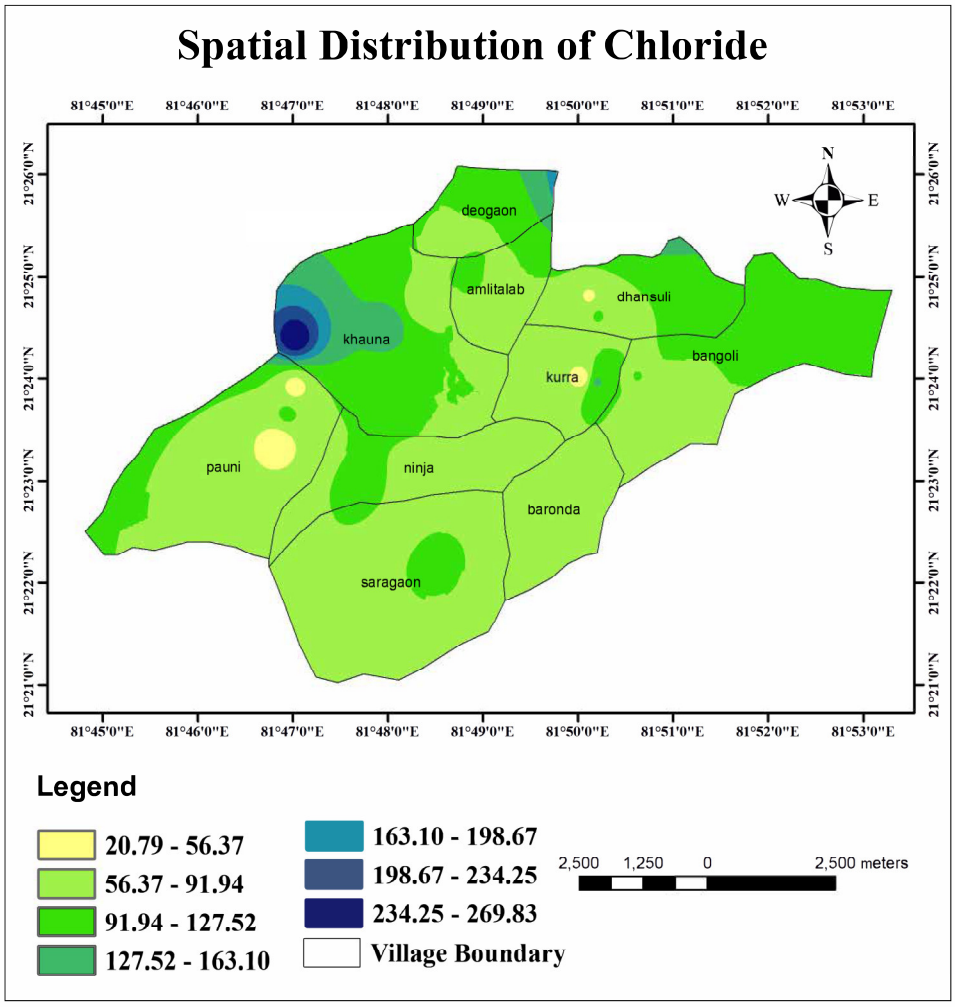


Figure A4. Spatial distribution of chloride.

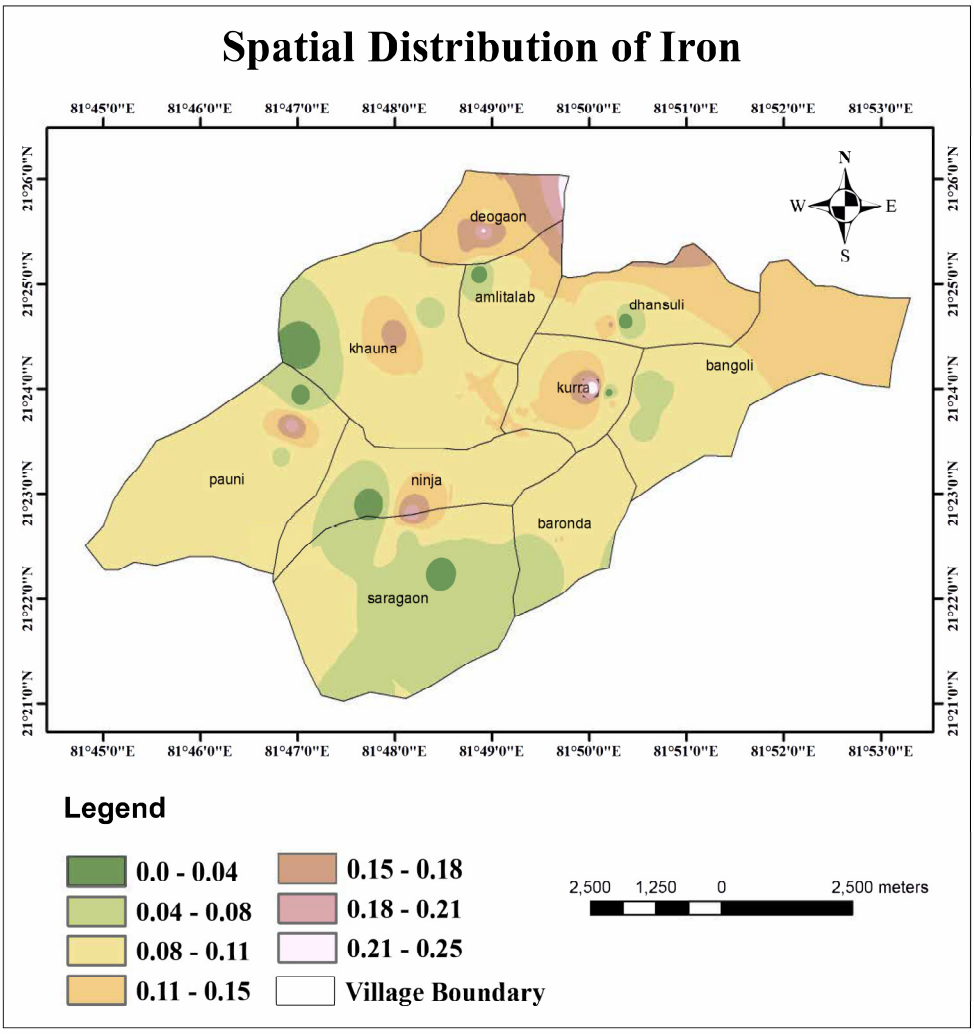


Figure A5. Spatial distribution of iron.

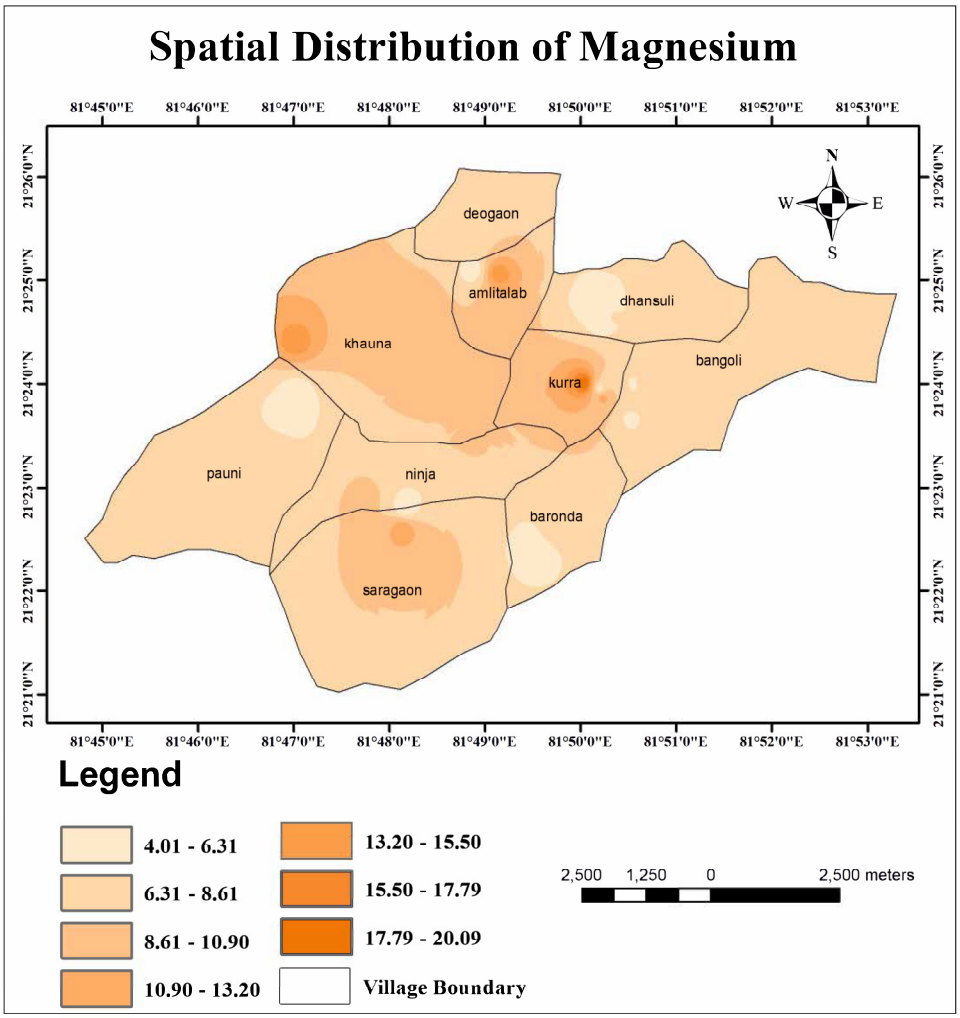


Figure A6. Spatial distribution of magnesium.

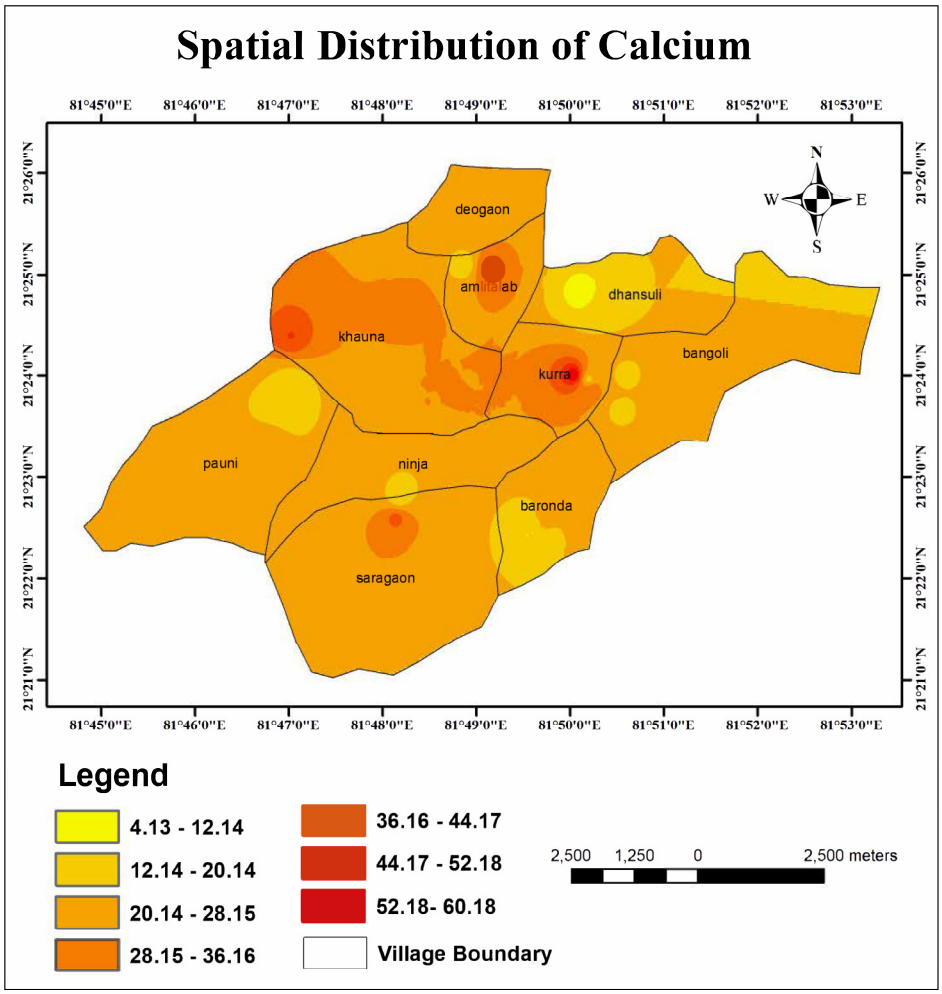


Figure A7. Spatial distribution of calcium.

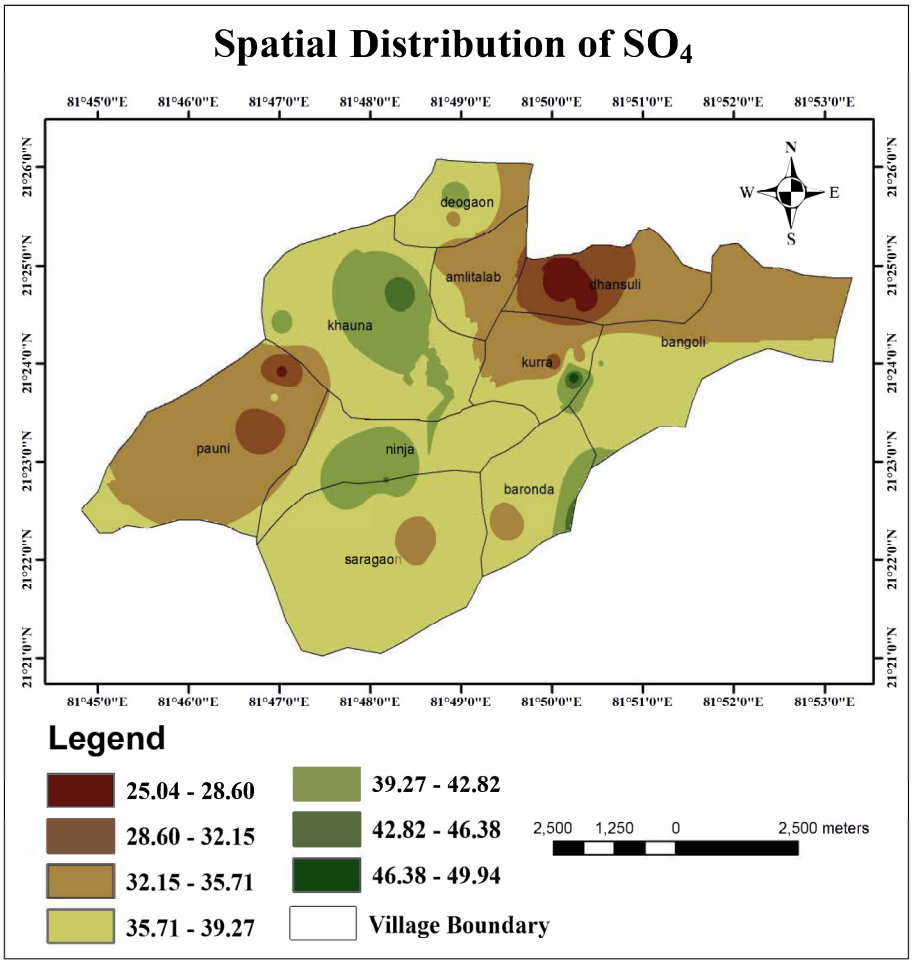


Figure A8. Spatial distribution of SO_4.

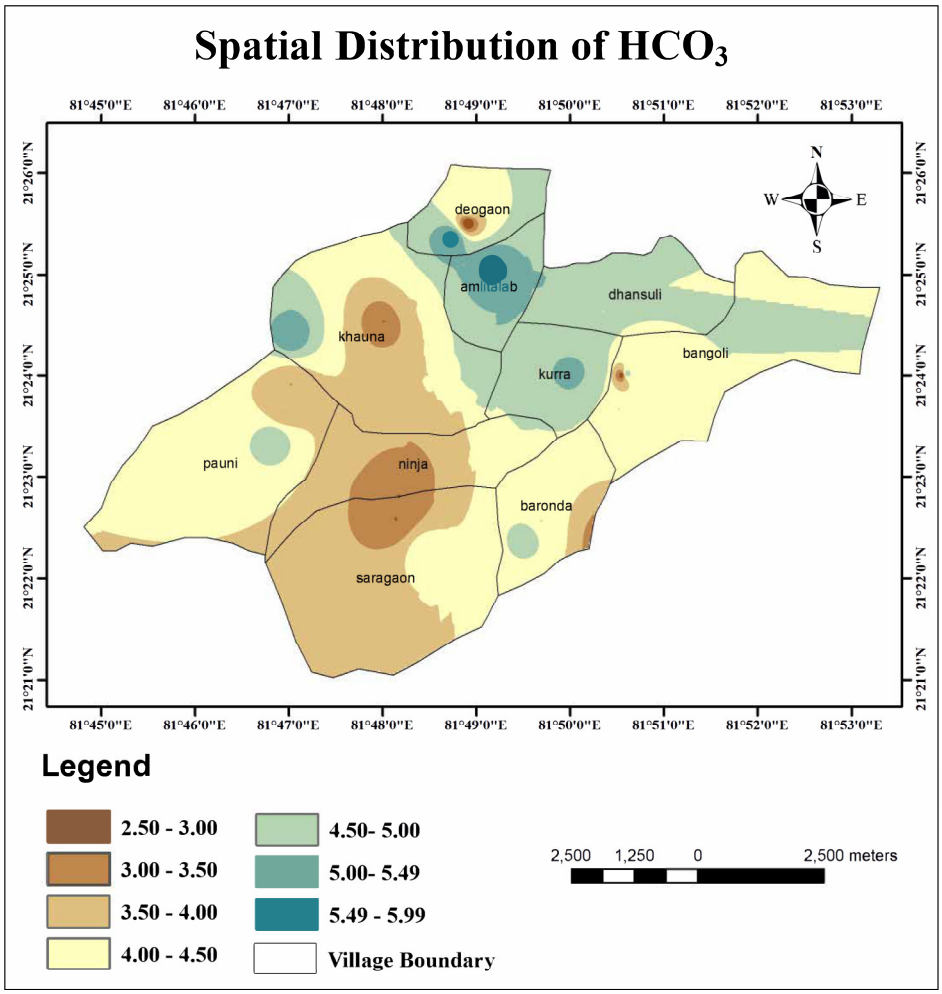


Figure A9. Spatial distribution of HCO_3.

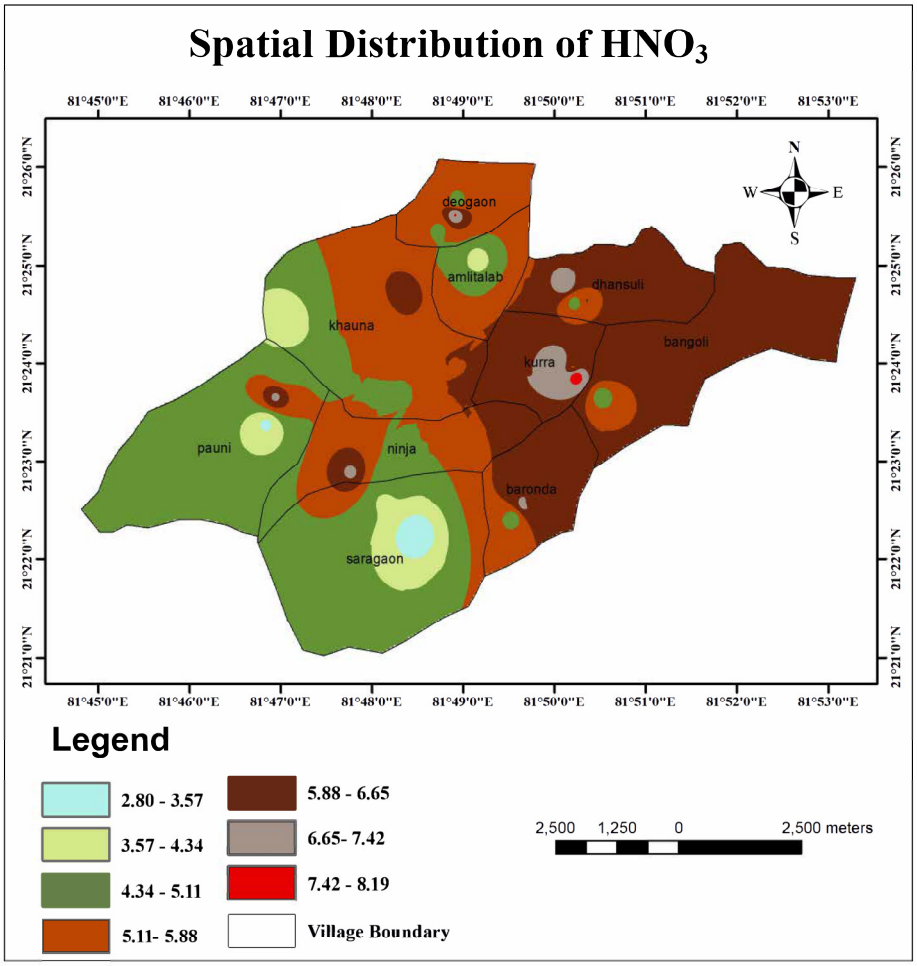


Figure A10. Spatial distribution of HNO_3.

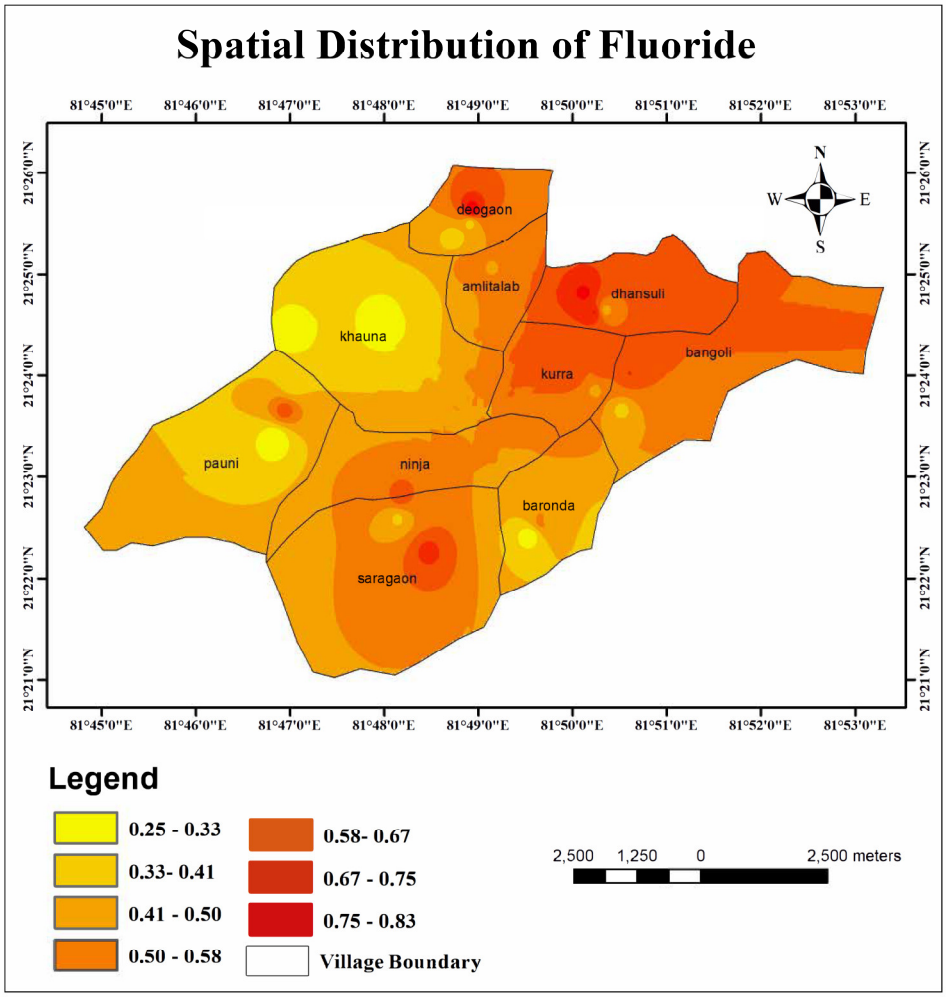


Figure A11. Spatial distribution of fluoride.

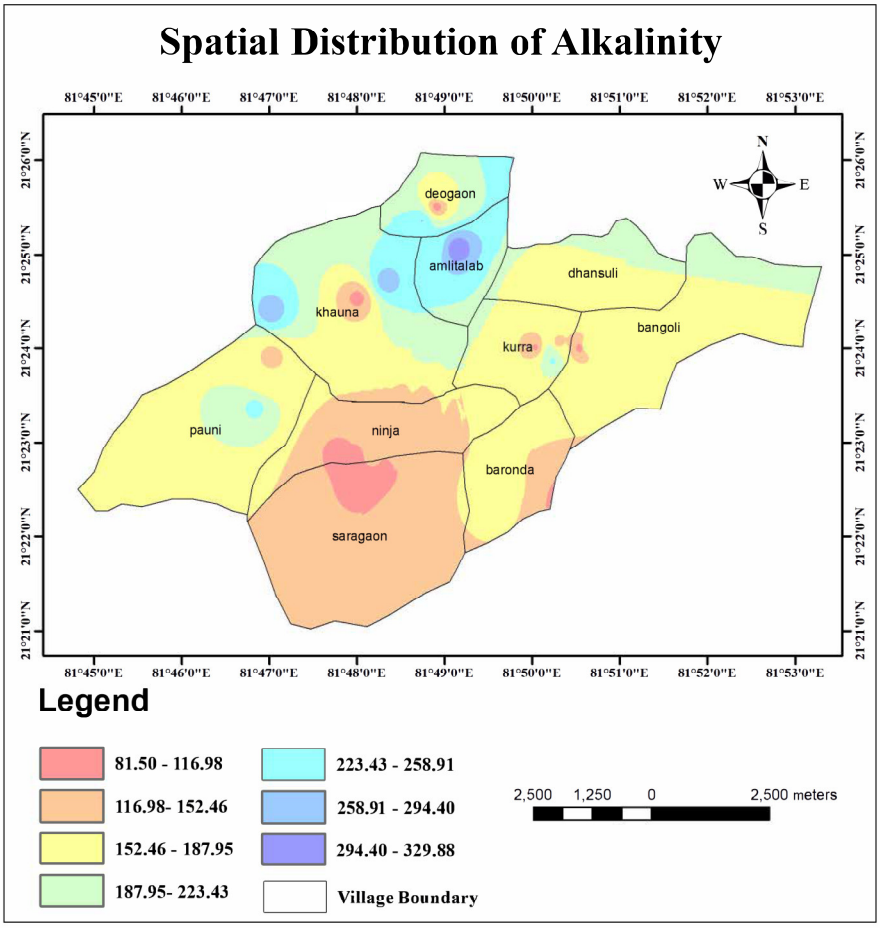


Figure A12. Spatial distribution of alkalinity.

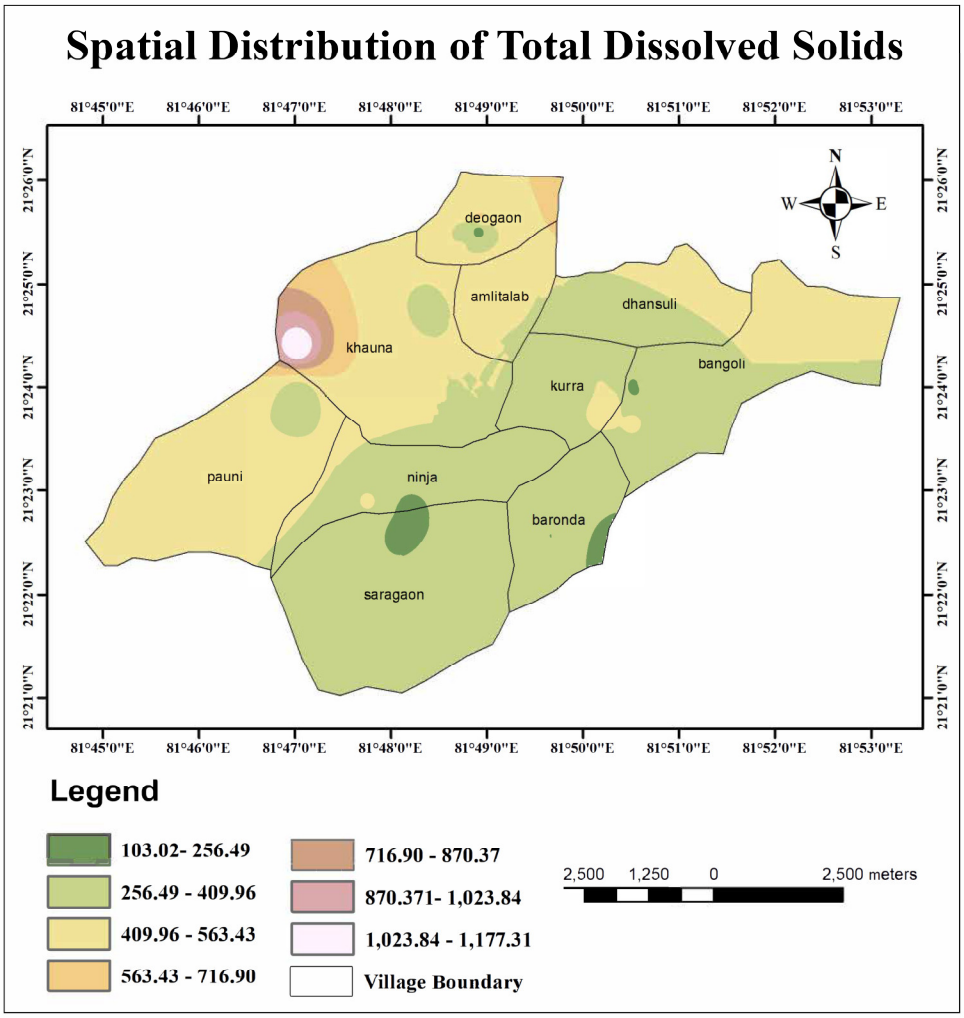


Figure A13. Spatial distribution of TDSs.

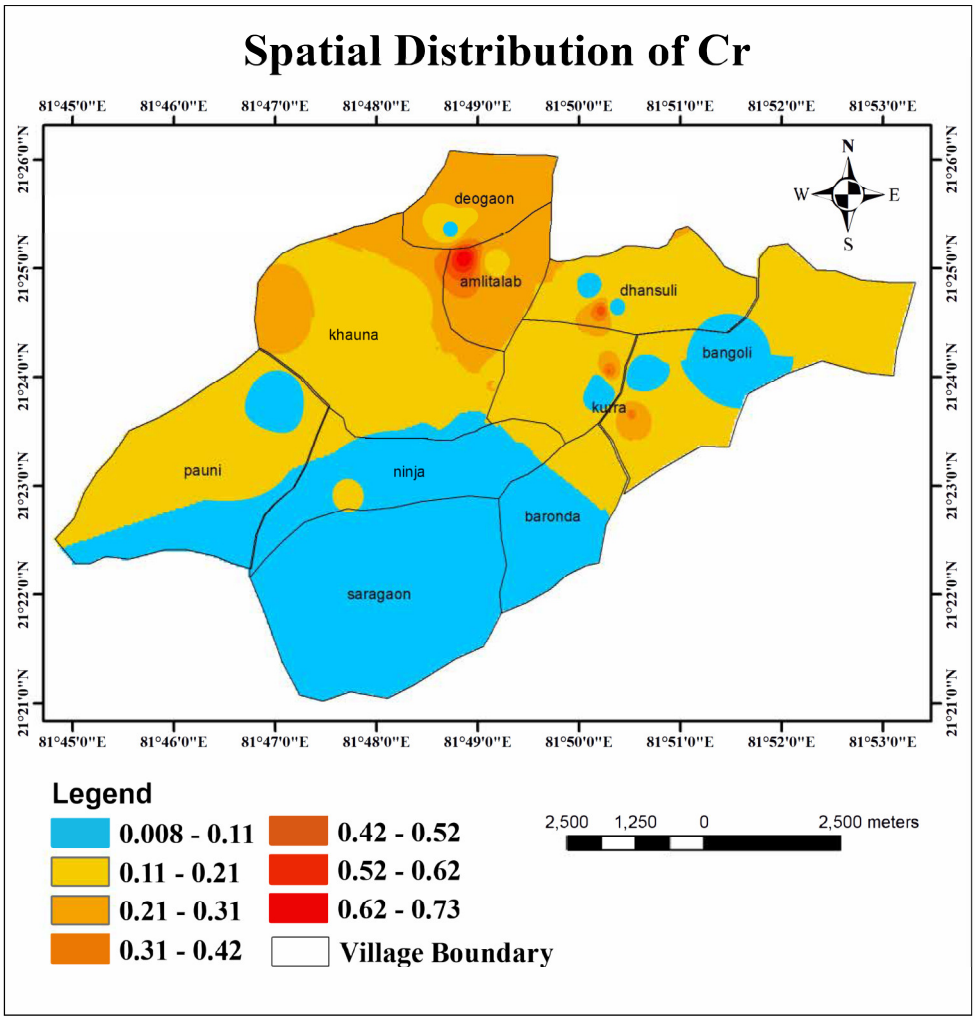


Figure A14. Spatial distribution of Cr.

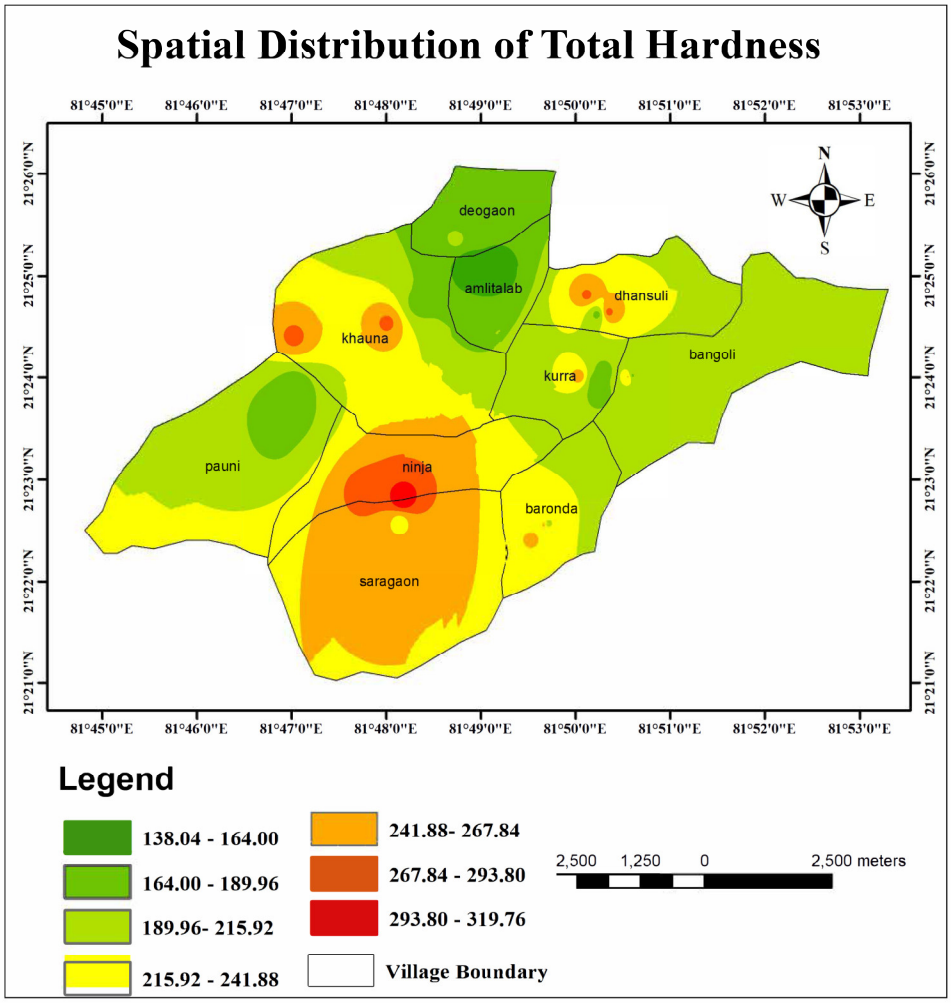


Figure A15. Spatial distribution of TH.

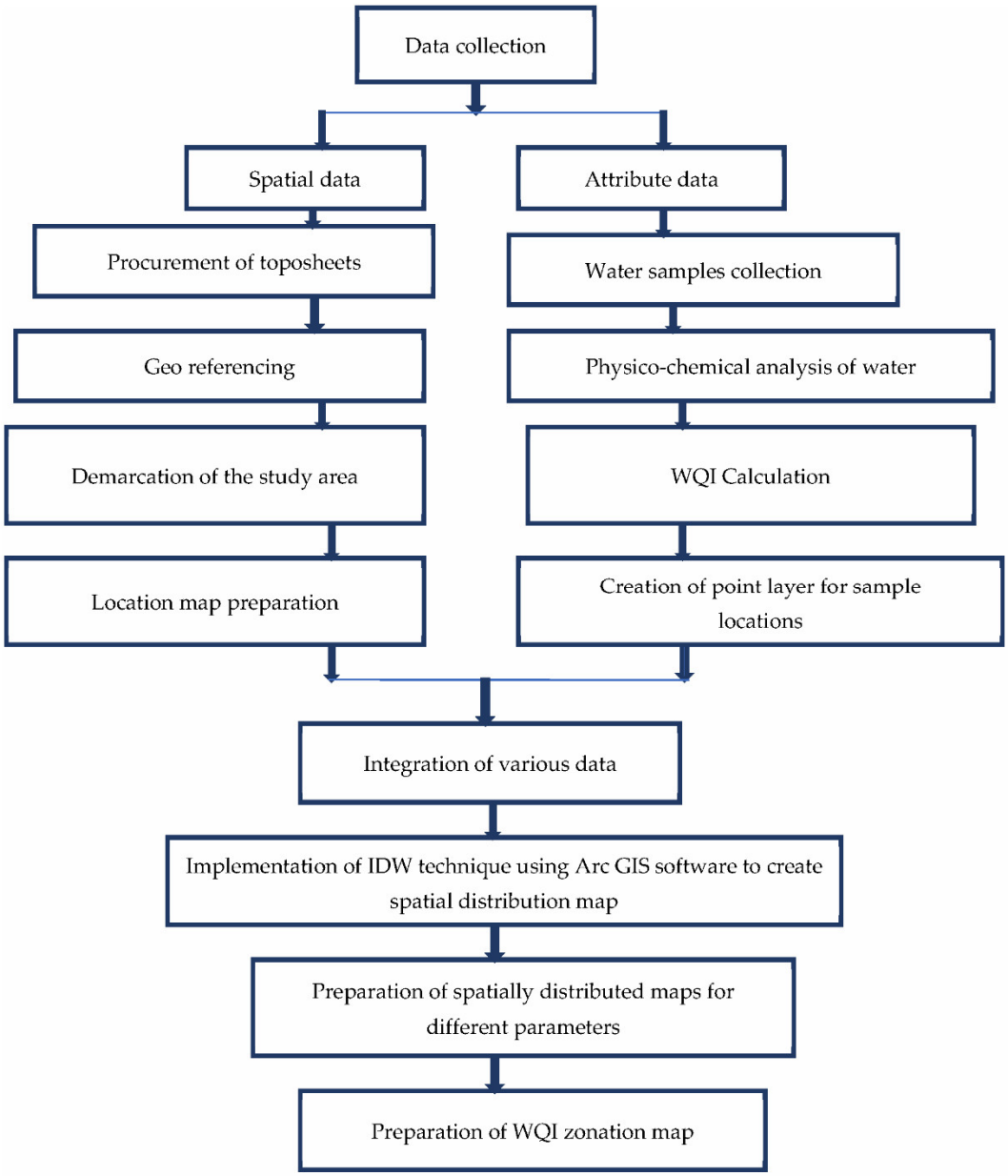


Figure A16. Flowchart of the procedure followed in the study.

References

1. Islam, S.; Rasul, T.; Bin Alam, J.; Haque, M.A. Evaluation of Water Quality of the Titas River Using NSF Water Quality Index. *J. Sci. Res.* **2010**, *3*, 151, doi:10.3329/jsr.v3i1.6170.
2. Al-Zahrani, M.A.; Abo-Monasar, A. Urban Residential Water Demand Prediction Based on Artificial Neural Networks and Time Series Models. *Water Resour. Manag.* **2015**, *29*, 3651–3662, doi:10.1007/s11269-015-1021-z.
3. Elkiran, G.; Ergil, M. The assessment of a water budget of North Cyprus. *Build. Environ.* **2006**, *41*, 1671–1677, doi:10.1016/j.buildenv.2005.06.014.
4. Shalby, A.; Elshemy, M.; Zeidan, B.A. Assessment of Climate Change Impacts on Water Quality Parameters of Lake Burullus, Egypt. *Environ. Sci. Poll. Res.* **2019**, *27*, 1–22.
5. Kavitha, R.; Elangovan, K. Ground water quality characteristics at Erode district, Tamilnadu India. *Int. J. Environ. Sci.* **2010**, *1*, 163–175.
6. Sharma, D.; Kansal, A. Water quality analysis of River Yamuna using water quality index in the national capital territory, India (2000–2009). *Appl. Water Sci.* **2011**, *1*, 147–157, doi:10.1007/s13201-011-0011-4.
7. Smith, D.G. A better water quality indexing system for rivers and streams. *Water Res.* **1990**, *24*, 1237–1244, doi:10.1016/0043-1354(90)90047-a.

8. Kannel, P.R.; Lee, S.; Lee, Y.-S.; Kanel, S.R.; Khan, S.P. Application of Water Quality Indices and Dissolved Oxygen as Indicators for River Water Classification and Urban Impact Assessment. *Environ. Monit. Assess.* **2007**, *132*, 93–110, doi:10.1007/s10661-006-9505-1.
9. Singh, R.P.; Nath, S.; Prasad, S.C.; Nema, A.K. Selection of Suitable Aggregation Function for Estimation of Aggregate Pollution Index for River Ganges in India. *J. Environ. Eng.* **2008**, *134*, 689–701, doi:10.1061/(asce)0733-9372(2008)134:8(689).
10. Yadav, N.S.; Kumar, A.; Sharma, M. Ecological health assessment of Chambal River using water quality parameters. *J. Integr. Sci. Technol.* **2014**, *2*, 52–56.
11. Agrawal, P.; Sinha, A.; Pasupuleti, S.; Nune, R.; Saha, S. Geospatial Analysis Coupled with Logarithmic Method for Water Quality Assessment in Part of Pindrawan Tank Command Area in Raipur District of Chhattisgarh. In *Climate Impacts on Water Resources in India*; Springer: Berlin, Germany, 2021; pp. 57–78.
12. Hameed, M.; Sharqi, S.S.; Yaseen, Z.M.; Afan, H.A.; Hussain, A.; Elshafie, A. Application of artificial intelligence (AI) techniques in water quality index prediction: A case study in tropical region, Malaysia. *Neural Comput. Appl.* **2017**, *28*, 893–905, doi:10.1007/s00521-016-2404-7.
13. Rufino, F.; Busico, G.; Cuoco, E.; Darrah, T.H.; Tedesco, D. Evaluating the Suitability of Urban Groundwater Resources for Drinking Water and Irrigation Purposes: An Integrated Approach in the Agro-Aversano Area of Southern Italy. *Environmental Monitoring and Assessment* **2019**, *191*, 1–17.
14. Vadiati, M.; Asghari-Moghaddam, A.; Nakhaei, M.; Adamowski, J.; Akbarzadeh, A. A fuzzy-logic based decision-making approach for identification of groundwater quality based on groundwater quality indices. *J. Environ. Manag.* **2016**, *184*, 255–270, doi:10.1016/j.jenvman.2016.09.082.
15. Bournaris, T.; Papathanasiou, J.; Manos, B.; Kazakis, N.; Voudouris, K. Support of irrigation water use and eco-friendly decision process in agricultural production planning. *Oper. Res.* **2015**, *15*, 289–306, doi:10.1007/s12351-015-0178-9.
16. Sargaonkar, A.; Deshpande, V. Development of an Overall Index of Pollution for Surface Water Based on a General Classification Scheme in Indian Context. *Environ. Monit. Assess.* **2003**, *89*, 43–67, doi:10.1023/a:1025886025137.
17. Gazzaz, N.M.; Yusoff, M.K.; Aris, A.Z.; Juahir, H.; Ramli, M.F. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar. Pollut. Bull.* **2012**, *64*, 2409–2420, doi:10.1016/j.marpolbul.2012.08.005.
18. Leong, W.C.; Bahadori, A.; Zhang, J.; Ahmad, Z. Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *Int. J. River Basin Manag.* **2019**, 1–8, doi:10.1080/15715124.2019.1628030.
19. Iticescu, C.; Georgescu, L.P.; Murariu, G.; Topa, C.; Timofti, M.; Pintilie, V.; Arseni, M.; Timofti, M. Lower Danube Water Quality Quantified through WQI and Multivariate Analysis. *Water* **2019**, *11*, 1305, doi:10.3390/w11061305.
20. Yaseen, Z.M.; Ramal, M.M.; Diop, L.; Jaafar, O.; Demir, V.; Kisi, O. Hybrid adaptive neuro-fuzzu models for water quality index estimation. *Water Resour. Manag.* **2018**, *32*, 2227–2245.
21. Diamantopoulou, M.J.; Papamichail, D.M.; Antonopoulos, V.Z. The use of a Neural Network technique for the prediction of water quality parameters. *Oper. Res.* **2005**, *5*, 115–125, doi:10.1007/bf02944165.
22. Khalil, B.; Ouarda, T.; St-Hilaire, A. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *J. Hydrol.* **2011**, *405*, 277–287, doi:10.1016/j.jhydrol.2011.05.024.
23. Gupta, R.; Singh, A.N.; Singhal, A. Application of ANN for Water Quality Index. *Int. J. Mach. Learn. Comput.* **2019**, *9*, 688–693, doi:10.18178/ijmlc.2019.9.5.859.
24. Isiyaka, H.A.; Mustapha, A.; Juahir, H.; Phil-Eze, P. Water quality modelling using artificial neural network and multivariate statistical techniques. *Model. Earth Syst. Environ.* **2019**, *5*, 583–593, doi:10.1007/s40808-018-0551-9.
25. Nourani, V.; Elkiran, G.; Abdullahi, J. Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J. Hydrol.* **2019**, *577*, 123958, doi:10.1016/j.jhydrol.2019.123958.
26. Gaya, M.S.; Abba, S.I.; Abdu, A.M.; Tukur, A.I.; Saleh, M.A.; Esmaili, P.; Wahab, N.A. Estimation of Water Quality Index Using Artificial Intelligence Approaches and Multi-Linear Regression. *Int. J. Artif. Intell. ISSN* **2020,** *2252*, 8938.
27. Najah, A.; El-Shafie, A.H.; Karim, O.A. Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environ. Sci. Pollut. Res.* **2014**, *21*, 1658–1670, doi:10.1007/s11356-013-2048-4.
28. Ahmed, A.N.; Othman, F.B.; Afan, H.A.; Ibrahim, R.K.; Fai, C.M.; Hossain, S.; Ehteram, M.; Elshafie, A. Machine learning methods for better water quality prediction. *J. Hydrol.* **2019**, *578*, 124084, doi:10.1016/j.jhydrol.2019.124084.
29. Karim, S.A.A.; Kamsani, N.F. Water Quality Index Using Fuzzy Regression. In *Water Quality Index Prediction Using Multiple Linear Fuzzy Regression Model*; Springer: New York, NY, USA, 2020; pp. 37–53.
30. Nayak, J.G.; Patil, L.; Patki, V.K. Development of water quality index for Godavari River (India) based on fuzzy inference system. *Groundw. Sustain. Dev.* **2020**, *10*, 100350, doi:10.1016/j.gsd.2020.100350.
31. Yasin, M.I.; Karim, S.A.A. A New Fuzzy Weighted Multivariate Regression to Predict Water Quality Index at Perak Rivers. In *Optimization Based Model Using Fuzzy and Other Statistical Techniques Towards Environmental Sustainability*; Springer: New York, NY, USA, 2020; pp. 1–27.
32. Abobakr Yahya, A.S.; Ahmed, A.N.; Binti Othman, F.; Ibrahim, R.K.; Afan, H.A.; El-Shafie, A.; Fai, C.M.; Hossain, M.S.; Ehteram, M.; Elshafie, A. Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios. *Water* **2019**, *11*, 1231.

33. Abba, S.I.; Pham, Q.B.; Saini, G.; Linh, N.T.T.; Ahmed, A.N.; Mohajane, M.; Khaledian, M.; Abdulkadir, R.A.; Bach, Q.-V. Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. *Environ. Sci. Pollut. Res.* **2020**, *27*, 41524–41539, doi:10.1007/s11356-020-09689-x.
34. Elkiran, G.; Nourani, V.; Abba, S. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J. Hydrol.* **2019**, *577*, 123962, doi:10.1016/j.jhydrol.2019.123962.
35. BIS (Bureau of Indian Standard). *Indian Standard Drinking Water–Specification*; Second Revision; Bureau of Indian Standards (BIS): New Delhi, India, 2012.
36. Chaurasia, A.K.; Pandey, H.K.; Tiwari, S.K.; Prakash, R.; Pandey, P.; Ram, A. Groundwater Quality assessment using Water Quality Index (WQI) in parts of Varanasi District, Uttar Pradesh, India. *J. Geol. Soc. India* **2018**, *92*, 76–82, doi:10.1007/s12594-018-0955-1.
37. WHO. *Guidelines for Drinking Water, Recommendations*; World Health Organization (WHO): Geneva, Switzerland, 2012.
38. Yisa, J.; Jimoh, T. Analytical studies on water quality index of river Landzu. *Am. J. Appl. Sci.* **2010**, *7*, 453.
39. Tyagi, S.; Singh, P.; Sharma, B.; Singh, R. Assessment of Water Quality for Drinking Purpose in District Pauri of Uttarakhand, India. *Appl. Ecol. Environ. Sci.* **2014**, *2*, 94–99.
40. Akter, T.; Jhohura, F.T.; Akter, F.; Chowdhury, T.R.; Mistry, S.K.; Dey, D.; Barua, M.K.; Islam, A.; Rahman, M. Water Quality Index for measuring drinking water quality in rural Bangladesh: A cross-sectional study. *J. Health Popul. Nutr.* **2016**, *35*, 1–12, doi:10.1186/s41043-016-0041-5.
41. Ramakrishnaiah, C.R.; Sadashivaiah, C.; Ranganna, G. Assessment of Water Quality Index for the Groundwater in Tumkur Taluk, Karnataka State, India. *E-J. Chem.* **2009**, *6*, 523–530, doi:10.1155/2009/757424.
42. 42. Eberhart, R.; Kennedy, J. *A New Optimizer Using Particle Swarm Theory*; IEEE: Piscataway, NJ, USA, 1995; pp. 39–43.
43. Gilani, S.-O.; Sattarvand, J.; Hajihassani, M.; Abdullah, S.S. A stochastic particle swarm based model for long term production planning of open pit mines considering the geological uncertainty. *Resour. Policy* **2020**, *68*, 101738, doi:10.1016/j.resourpol.2020.101738.
44. Yasin, Q.; Sohail, G.M.; Ding, Y.; Ismail, A.; Du, Q. Estimation of Petrophysical Parameters from Seismic Inversion by Combining Particle Swarm Optimization and Multilayer Linear Calculator. *Nat. Resour. Res.* **2020**, *29*, 3291–3317, doi:10.1007/s11053-020-09641-3.
45. Mehrabi, M.; Pradhan, B.; Moayedi, H.; Alamri, A. Optimizing an Adaptive Neuro-Fuzzy Inference System for Spatial Prediction of Landslide Susceptibility Using Four State-of-the-art Metaheuristic Techniques. *Sensors* **2020**, *20*, 1723, doi:10.3390/s20061723.
46. Sun, D.; Wen, H.; Wang, D.; Xu, J. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphol.* **2020**, *362*, 107201, doi:10.1016/j.geomorph.2020.107201.
47. Gafar, A.A.; Khayat, M.E.; Ahmad, S.A.; Yasid, N.A.; Shukor, M.Y. Response Surface Methodology for the Optimization of Keratinase Production in Culture Medium Containing Feathers by *Bacillus* sp. UPM-AAG1. *Catalysts* **2020**, *10*, 848, doi:10.3390/catal10080848.
48. Bui, Q.-T.; Nguyen, Q.-H.; Nguyen, X.L.; Pham, V.D.; Nguyen, H.D.; Pham, V.-M. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *J. Hydrol.* **2020**, *581*, 124379, doi:10.1016/j.jhydrol.2019.124379.
49. Pourghasemi, H.R.; Razavi-Termeh, S.V.; Kariminejad, N.; Hong, H.; Chen, W. An assessment of metaheuristic approaches for flood assessment. *J. Hydrol.* **2020**, *582*, 124536, doi:10.1016/j.jhydrol.2019.124536.
50. Roshanravan, B.; Aghajani, H.; Yousefi, M.; Kreuzer, O. Particle Swarm Optimization Algorithm for Neuro-Fuzzy Prospectivity Analysis Using Continuously Weighted Spatial Exploration Data. *Nat. Resour. Res.* **2018**, *28*, 309–325, doi:10.1007/s11053-018-9385-4.
51. Engelbrecht, A.P. *Computational Intelligence: An Introduction*, 2nd ed.; Jonh Wiley & Sons, Ltd.: Chichester, UK, 2007; Volume 1.
52. Clerc, M.; Kennedy, J. The particle swarm—Explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* **2002**, *6*, 58–73, doi:10.1109/4235.985692.
53. Ciarelli, P.M.; Krohling, R.A.; Oliveira, E. Particle swarm optimization applied to parameters learning of probabilistic neural networks for classification of economic activities. In *Particle Swarm Optimization*; InTech: Rijeka, Croatia, 2009; pp. 313–327.
54. Scholkopf, B.; Sung, K.-K.; Burges, C.J.C.; Girosi, F.; Niyogi, P.; Poggio, T.; Vapnik, V. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Process.* **1997**, *45*, 2758–2765, doi:10.1109/78.650102.
55. Dibike, Y.B.; Velickov, S.; Solomatine, D.; Abbott, M.B. Model Induction with Support Vector Machines: Introduction and Applications. *J. Comput. Civ. Eng.* **2001**, *15*, 208–216, doi:10.1061/(asce)0887-3801(2001)15:3(208).
56. Li, C.-H.; Lin, C.-T.; Kuo, B.-C.; Ho, H.-H. An Automatic Method for Selecting the Parameter of the Normalized Kernel Function to Support Vector Machines. In Proceedings of the 2010 International Conference on Technologies and Applications of Artificial Intelligence, Hsinchu, Taiwan, 18–20 November 2010; pp. 226–232.
57. Shukla, R.; Chakraborty, A.; Sachdeva, K.; Joshi, P.K. Agriculture in the western Himalayas—An asset turning into a liability. *Dev. Pract.* **2017**, *28*, 318–324.
58. Shukla, R.; Sachdeva, K.; Joshi, P.K. Demystifying vulnerability assessment of agriculture communities in the Himalayas: a systematic review. *Nat. Hazards* **2017**, *91*, 409–429, doi:10.1007/s11069-017-3120-z.