Indian Journal of Science and Technology, Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/101360, December 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Algorithm Design for GIS using BFO and Neural Network Approach for Identification of Groundwater

Amit Verma*, **Jasleen Kaur and Bharti Chhabra**

Computer Science and Engineering, Chandigarh Engineering College, Landran - 140307, Punjab, India; Dramitverma.cu@gmail.com, kjasleen14@gmail.com, cecm.cse.bharti@gmail.com

Abstract

Objectives: Groundwater is an important resource contributing significantly in development of natural life. However, over exploitation has depleted groundwater ratio deliberately and also led to land subsidence at some places. **Methods/Statistical Analysis**: Groundwater zones are demarcated using remote sensing and Geographic Information System (GIS) techniques. **Findings:** In this research a definitive methodology is proposed to determine groundwater using integration of BFO and neural network technique. For the training purpose we use fuzzy logic and after that we use optimization techniques to find suitable feature set which can classify more accurate. **Applications/Improvements:** Finally, it is concluded that the Geoinformatics technology are very efficient and useful for the identification of groundwater detection. We evaluate parameters like kappa coefficient, water level, accuracy of algorithm to detect the water percentage in the region from where we conduct the satellite image through remote sensing. Therefore, this research will be useful for effectual identification of suitable locations for extraction of water.

Keywords: BFO, Fuzzy Logic, Ground Water Detection, Kappa Coefficient. Neural Network

1. Introduction

Groundwater is a valuable and most broadly disseminated asset of the earth[1]. It constitutes a critical wellspring of water supply for different purposes, for example, local, mechanical and agrarian requirements[2].

Reduction in the groundwater level due to the increasing demands of the needs leads to inspect the course of events of groundwater exhaustion[3]. Water necessity of paddy is much higher than that of option products[4]. While the zone under rice harvest was under 5% of the gross trimmed range in India amid the mid 1970s, the extent expanded to more than 35% in the late times.

Amid most recent couple of decades, the greater part of the ranchers has moved from diffusive to submersible pumps[4]. Such a marvel of exhausting groundwater and consequent developing of borewells and moving to submersible pumps required huge speculations from the agriculturists furthermore prompted noteworthy expansion in ranchers' obligation[5]. The wonder of extending of borewells and ensuing movement from radial to submersible pumps was archived on the premise of ranchers' reactions. The ranchers were requested that report the year when their electric engines were most recently seen working at the surface before developing of the borewells began[6].

As medium and expansive ranchers possessed more than one engine, 45% and 6% of them were selecting to share the second and third engines individually. Half of the little ranchers and 30% of the medium agriculturist are referred to high speculations and littler landholdings as essential explanations behind their selecting to share the electric engines. Share method of electric engines to draw groundwater for watering system. Decrease in access to groundwater may directly affect the trimming design and may compel the ranchers to settle on less water-devouring products. The greater part of the progressions in editing design happened amid the kharif season and not amid the rabi season[7]. The progressions, in any case, show

*Author for correspondence

that there were endeavors to adjust to the changing access to groundwater assets. Aside from the progressions in editing design, there are some other adjustments choices, which can possibly bring about a critical decrease in the water, utilize and may help in countering the unfriendly effect of decrease in access to groundwater assets.

Remote detecting and GIS now-a-days have gotten to be inescapable devices for the investigation of groundwater at nearby, provincial and worldwide level[8]. Neural system based ideas have found an extensive variety of uses in various fields viz. soil science, ecological science, earth science and so forth. It gives an exceptionally exact way to deal with managing instability which becomes out of the unpredictability of human conduct.

So, in this research, hybrid algorithm is proposed to detect the ground water level in soil e.g. fuzzy logic is used to get the input, then obtained features are optimized using BFO. In the end, classification is done using neural network. Data mining concept is used for data collection, data creation and data management and for its analysis in this research[9]. The basic procedure of data mining is to extract knowledge from data in the context of large databases[10]. BFO (Bacterial Foraging Optimization Algorithm) is used for information processing strategy. Basically BFO is inspired by the behavior of bacteria to get optimal solutions. Neural network is utilized for testing purpose based on assigned weight functions. The basic aim of neural network is to work like human brain works. Neural network consists of various no. of neurons and their working is similar to the brain neuron structure.

2. Literature Survey

In this era, researchers have used number of techniques for detecting ground water percentage. Existing work for detecting percentage value of water by authors is presented here[11] discussed new algorithm, namely accelerated Fast global k-means it is a modified version of fast global k-means. Accelerated fast global k-means firstly apply local geometrical information. On the basis of that, an acceleration mechanism is developed that is used for the production of new cluster centers. FGKM+A mechanism needs less computing time and lesser distance calculations while retaining the same clustering results as that of the fast global k-means algorithm. The performance of the proposed algorithm increases as the number of clusters of a data set increases [12]clustering technique k-mean which is partition-based clustering method. K-mean firstly initializes the center and then calculates the distance of another element. K-mean creates the cluster of those elements whose distance is very less from the center.

In[13] discussed the modified version of k-mean algorithm for formulating clusters. Modified k-mean algorithm can perform clustering very effectively on Categorical data sets. [14]Proposed k-means algorithms as k-methods is less popular algorithm for clustering because it's too complex and costly than that of k-mean. The proposed algorithm describes the deviation which shortens the total time to formulate the cluster by simple k-mean.

In[15] proposed the classification of documents written in Arabic language using Artificial Neural Network (ANN). This technique has been used in limited version in previous years. Arabic documents have been collected from Arabic text corpus. Firstly training and then testing has been done using ANN classifier. In addition to this, weighted scheme for feature extraction has also been used. In[16] discussed modifying Genetic Algorithm that partition the mixed numeric and categorical data into clusters by finding a globally optimal partition. To handle the categorical data proposed method involves k-means algorithm with enhanced cost function, it is effective in recovering the elementary cluster structures. [17]Proposed a clustering based method named Hybrid Genetic K-means Algorithm (HGKA) that is applicable only for numeric data set

3. Basic Concepts

This section describes the basic concepts which is used for ground water detection is Fuzzy Logic, Bacterial Foraging Optimization and Neural Network Back Propagation algorithm.

3.1 Fuzzy Logic

Fuzzy logic is a good training algorithm in which training is done using membership functions. It basic four main steps are:

- **Fuzzification Module** – It transforms the system inputs, which are crisp numbers, into fuzzy sets.
- **Knowledge Base –** It store IF THEN rules.
- **Inference Engine** – It stimulates the human reasoning process by making the fuzzy inference on the values and if then rules.
- **Defuzzification Module** – It transforms the fuzzy set obtained by the inference engine into a crisp value.

3.2 Bacterial Foraging Optimization

The Bacterial Foraging Optimization Algorithm is proposed by Kevin Passino inspired by the group foraging strategy of swarm of bacteria such as E. coli and M. xanthus[18]. During Foraging two basic operations performed by bacterium are:

Tumble: Clockwise Rotation (CW): CW causing the bacterium to tumble in place, breaks the flagella bundles apart, such that they start moving in different directions.

Swim: Counter Clockwise Rotation (CCW)[19]: CCW causing the bacterium to swim in straight line, aligns the flagella into a single rotating bundle.

Some bacteria such as E. coli have several flagella per cell (4-10 typically). These can rotate in two ways:

3.3 Neural Network

NN is the most commonly used classification algorithm and it works on two steps;

In training section, classifier learns its own classification rules[20].

In testing phase, the feature vectors of reduced clusters with reduced Bfo features takes as input and resultant parameters generated with zero error and definite accuracy.

4. Methodology

In this section, Extended BFO is implemented with a new fitness function that detects the ground water quantity in particular that depends on various features extracted. Kappa coefficient parameter is used for detecting the ground water possibility.

Lemma 1: Feature Extraction Algorithm

```
Step 1: Load ('maindata','data');
 {
                    // Loading the database from maindata
Step 2: data11=data(1:50,1:3);
{
                                  // Range for row to column
Step 3: data11=data;
          }
//for loop
                    {
[data1,data2,i1,i2,loc1,loc2]=fuzzyruleset(data11);
                                    //Arguments
axes(handles.axes1);
          // Nomenclature for vertical and horizontal axis
                    }
//For loop
                              {
Step 4: % for i=1:numel(data1)
          // Initialization of loop for data set 1
                              }
Step 5: plot(loc1.x, loc1.y, 'ro');
                            // Plotting value for said variable
 hold on;
 text(loc1.x+2, loc1.y+2, 'C1');
xlabel('GPS X');
ylabel('GPS Y');
                                                  // Labeling
End
 }
```

Description: Above Algorithm is for features extraction of uploaded dataset. First, we upload the data set which is in the form of excel sheet. Secondly, we extract the data values from data set and apply fuzzy rule set to extract the features of satellite data. Extracted features are stored in the clusters.

Lemma 2: Optimization Algorithm

```
Step 1: total bact1=numel(data1);
totalbact2=numel(data2);
          // calculate the total bacteria in data set1
     {
Step 2: reducedindexcount=0;
{
Bforeduced1=[];
                              //BFO reduced count
for i=1:totalbact1
      {
Current swim=data1(i);
          // BFO operaration swim (CCW) on data 1
f=bfo fitness( currentswim,data1);
                    // calculate Fitness function (f_s)
                    {
if f==1
reduced indexcount=reducedindexcount+1;
Bfo reduced1(reducedindexcount)=currentswim;
 End
         }
   End
       }
}
```

```
Reduced index count=0;
{
Bfo reduced2=[];
                    //BFO reduced count for data 2
For i=1:totalbact2
          {
Current swim=data2(i);
          // BFO operaration swim (CCW) on data 2
f=bfofitness( currentswim ,data2);
                    // calculate Fitness function (f_s)
           {
if f==1
reducedindexcount=reducedindexcount+1;
Bforeduced2(reducedindexcount)=currentswim;
 End
           }
  End
        }
}
```

Description: Above Algorithm for features reduction of the uploaded dataset after the application of BFO algorithm. First, we have to find the total count of bacteria. Secondly,We calculate the fitness function to reduced the features of clustered data. We operate swim operation on data to find the fitness function. At last reduce index count describe the reduced features of data set.

Lemma 3: Classification Algorithm

```
Step 1: loadbfofeatures
                              //load the reduced features
{
Step 2: s1g=numel(Bforeduced1);
   {
trainingpercentage1=round(s1g*70/100);
trainingdata   (1,1:trainingpercentage1)=Bforeduced1(1,
1:trainingpercentage1);
   // Train the 70% data from the BFO reduced data1
   }
s2g=numel(Bforeduced2);
   {
trainingpercentage2=round(s2g*70/100);
trainingdata(2,1:trainingpercentage2)=Bforeduced2(1,
1:trainingpercentage2);
// Train the 70% data from the BFO reduced data2
   }
Step 3: group(1)=1;
Step 4: group(2)=2;
                              // group formation
Step 5: net=newff(trainingdata',group,10);
net.trainParam.epochs=100;
                    // evaluate the training parameters
net=train(net,trainingdata',group);
}
```

Description: Above algorithm is for the classification of features extracted using optimization algorithm. First, we load the BFO reduced features from database. Then we calculate the training percentage of both the data set generated after optimization[21]. Training parameters are evaluated for reduced features to find the kappa coefficient.

Lemma 4: Accuracy Algorithm

```
Step1: leftins1=s1g-trainingpercentage1;
leftins2=s2g-trainingpercentage2;
{
          //calculate the remaining data for testing
Step2: test_Data(1,:)=Bforeduced1(1,410:430);
test_Data(2,:)=Bforeduced2(1,220:240);
// if optimization algorithm is correct then test data is
equal to reduced feature values
Step 3: [r, c]=size(test_Data);
                    // calculate the rows amd column
          {
test_Data(1,c:303)=test_Data(1,c);
test_Data(2,c:303)=test_Data(2,c);
      }
Step 4: result=sim(net, test_Data');
error=abs((mean(result)-mean(mean(test_Data)))/
numel(test_Data));
          // check the error occur during classification
accuracy=100-error;
// if no error will occur then classification algorithm is
accurate
}
```

Description: Above mentioned algorithm is used for finding the accuracy of features classified or tested using neural network. First we Calculate the remaining data for testing i.e. 30%. If the optimization algorithm is accurate then the teat data will be equal to the reduced feature values. Calculate the rows and column of test data and check the error. If no error occur then BFO and Neural Network Techniques which we are applying are accurate.

Lemma 5: Kappa Coefficient

```
Step1: kappa_coef1=(numel(Bforeduced1)/numel(test_
       Data))*(r/(numel(Bforeduced1)/c)+1);
```

kappa_coef2=(numel(Bforeduced2)/numel(test_Data))*(r/(numel(Bforeduced2)/c)+1);
//calculate the kappa coefficient for both reduced datasets
Step 2: kappa_coef = (kappa_coef1+kappa_coef2)/2;
//calculate mean kappa coefficient
Step 3: waterdetectionpercentage = kappa_coef*100;
// calculate water percentage

Above mentioned algorithm is used to describe the water level percentage in the regions trained during clustering. First we calculate the kappa coefficient of both the reduced data sets. Then calculate the mean kappa coefficient and at last evaluate the water percentage.

The complete data flow for the implemented concept are shown in the Figures 1 to 3 in form of Context DFD, Level 0 and Level 1 DFD.

Figure 1 shows the Context level data flow diagram describe the simple view of research. It describes the satellite data uploaded from repository to ground water prediction process which perform the operations on the data set and calculate the end parameters through classification algorithm.

Figure 2 shows the Level 1 diagram describe the processing of Ground water prediction process. This process determines the way to evaluate the parameters to find the water level. In this methodology first we define the clustering by using fuzzy logic. After clusters are generated based on feature extraction mechanism we apply Bacterial Foraging optimization algorithm to reduced clusters and then we apply classification algorithm to train and test

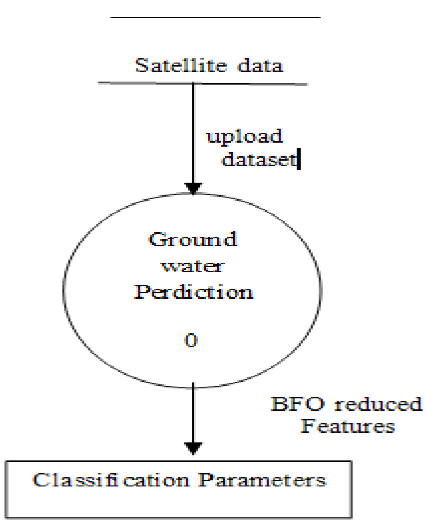

Figure 1. Context level diagram.

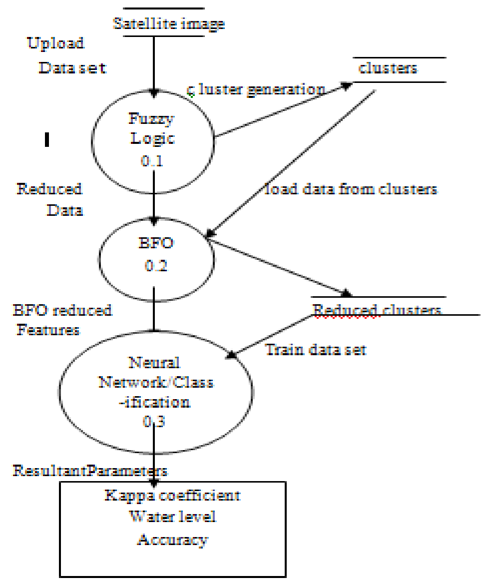

Figure 2. Level 1 data flow diagram.

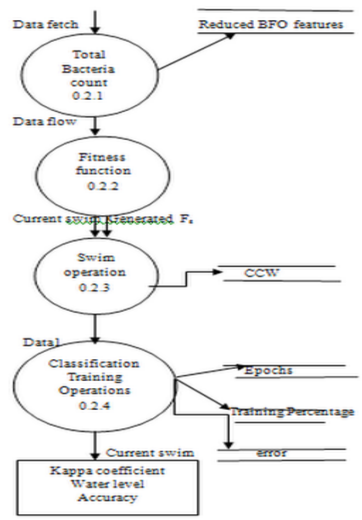

Figure 3. Level 2 data flow diagram.

Figure 3 shows the Level 2 Diagram describe the detailed processing of Bacterial Foraging Optimization algorithm. Foraging operations are of two kinds the bacteria posses. In are research bacteria performs swim operation to calculate the fitness function (F_s). Fitness function has two parameters one is current swim and data in clusters. Fitness Function generated which helps in find the healthy bacteria count and training percentage. After computing training data, we perform classification operations to evaluate the epochs. When neural network perform operations the then resultant parameters get generated. These are Kappa coefficient, water level, accuracy. These parameters help in prediction of water in region.

5. Experimental Results and Discussion

This section evaluates the performance of Fuzzy Logic, Optimization Algorithm i.e. BFO, Classification algorithm Neural Network. Windows 7 based system with 4 GB of RAM, 500 GB of HDD, an Intel(R) Core(TM) i7 CPU, is used for conducting this research. Simulation tool MatlabR2010a is used is used for implementation. For overall evaluation of implemented concept, parameters of kappa coefficient, water percentage and accuracy are considered.

5.1 Evaluation Parameters

There are various available metrics used to evaluate different techniques. In current work, performance parameters of kappa Coefficient, Water Percentage and Accuracy have been evaluated. .

5.2 Kappa Coefficient

With the varied number of variations, kappa coefficient values are shown. The value for this parameter is started from 1.6576 and by the number of iterations; the value is decreased to 1.5435. The percentage of the values in the major diagonal of the table is measured by Kappa. Table 1 shows the table for the kappa coefficient with the varied number of values. Figure 4 shows the graph with variation in values of kappa coefficient.

Table 1. Kappa coefficient

S. No.	Proposed Work
1	1.6576
2	1.6853
3	1.2435
4	1.8776
5	1.5435

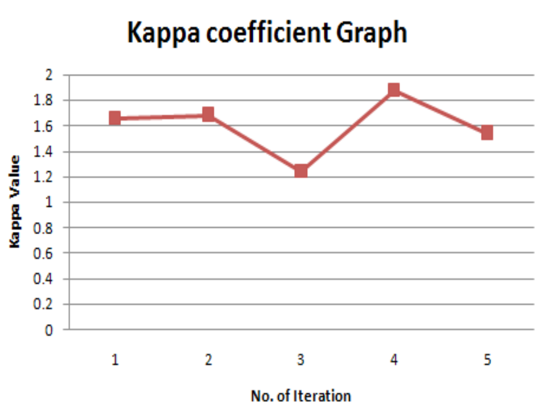

Figure 4. Kappa coefficient graph.

5.3 Water Level Percentage

With the varied number of variations in kappa coefficient, water level percentage is calculated. The values are varied from 65.76 and decreased up to 62.54. The values for this parameter are varied simultaneously. Table 2 is for the water level with respect to the number of iterations. Figure 5 shows the water level in the graph is decreased.

5.4 Accuracy

With the amount of iterations, rate of accuracy is varying. Accuracy parameter shows the reliability of the work done. With 99.34% accuracy, it is said the work is 99.34% reliable. Figure 6 shows the parameter evaluation being done using three basic parameters i.e. kappa = 1.65, accuracy 99% and water level percentage = 65.92%. As per

Table 2. Water Level (%)

S. No.	Proposed Work
1	65.76
2	67.34
3	68.34
4	63.49
5	62.54

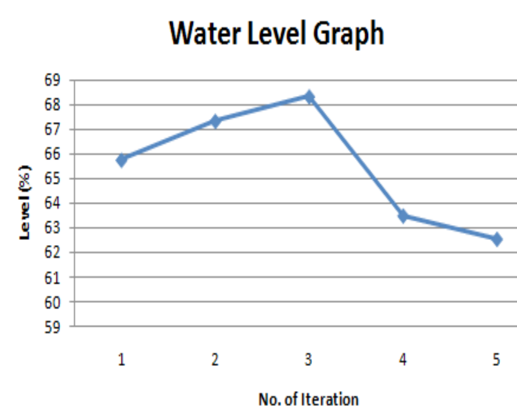

Figure 5. Water level % graph.

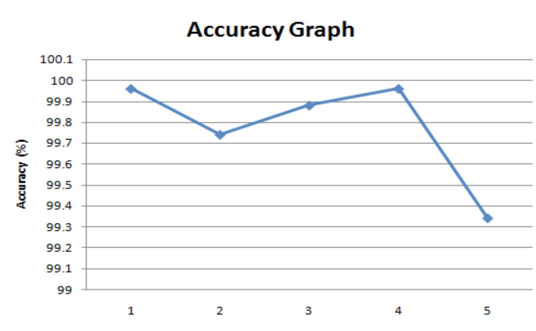

Figure 6. Accuracy graph.

Table 3. Accuracy

S. No.	Proposed Work
1	99.96
2	99.74
3	99.88
4	99.96
5	99.34

Table 3 (shows) one can check the accuracy parameter with the number of iterations.

The Kappa Coefficient (as per Equation 1) can be defined as the discrete multivariate technique that is used to understand the consequences of error matrix. The Kappa statistic takes both the off diagonal observations having rows and columns and the diagonal observations to provide an additional strong declaration of accuracy assessment as compare to the accuracy measures. The Kappa coefficient is measured by using the subsequent formula to the error matrix:

$$\hat{k} = \frac{N\sum_{i=1}^{r} y_{ii} - \sum_{i=1}^{r}(y_{i+}\cdot\cdot y)}{M^{2} - \sum_{i=1}^{r}(y_{i+}\cdot\cdot y_{+i})}$$

Equation1

6. Conclusion

Due to overuse of water resources in agricultural land, ground water resources has become on the merge of the distinct. Therefore proper management is required for the management of ground water in ground water counter. One of the most central factors in flourishing recharge of groundwater possessions is locate fitting areas for this project. Therefore, site miscellany of false renew appropriate areas is very significant. Four factors namely, geology, land, soil and landform parameter were explore, secret, prejudiced and overlay using future organization.

To superimpose these features, CLUSTERING, BFO ALGORITHM and neural network were used. Results show that based on proposed method, obtained values of kappa = 1.65, accuracy 99% and water level percentage = 65.92%.

For the automation and robust clustering method one can implement adaboost with the optimization techniques. However for the large datasets one can implement cascading techniques as below explained in the algorithm.

Algorithm _ Optimization Improved

For the Given dataset 1 and 2

// the said algorithm will do the cascading and optimizing and the process of normalization will become less complex.

Step1: totalbact1=numel(data1);

totalbact2=numel(data2);

// calculate the total bacteria in data set1

{

Step 2: reducedindexcount=0;

{

Bforeduced1=[];

//BFO reduced count

for i=1:totalbact1

{

currentswim=data1(i);

// BFO operaration swim (CCW) on data 1

f=bfofitness(currentswim,data1);

// calculate Fitness function (f_s)

{

if f==1

reducedindexcount=reducedindexcount+1;

Bforeduced1(reducedindexcount)=currentswim;

End

}

End

}

}

reducedindexcount=0;

{

Bforeduced2=[];

//BFO reduced count for data 2

for i=1:totalbact2

{

currentswim=data2(i);

// BFO operation swim (CCW) on data 2

For all data2(i)

FunctionFunction(Cascaded parameter data2$(i)_n$)

Initialize $f_{positive}$, $d_{minimum}$

// The maximum acceptable false (positive) / total layer and d minimum acceptable detection rates / total layers

If

*/

while data2$(i)_n$ > Ftarget

– i←i + 1 – ni = 0;

F_i – F_i – 1

– while F_i > f x F_i –1

*ni ← ni + 1

*/

```
{
Fi>Ftarget. // where P and N
Positive and Negative example
Update weights
Evaluate the current values of
C1(x)
        // where c1(x) will be the
        cascading of all data set
      }
Else check the values of Fi
        //Where i=1 to n where n
        is number of parameters
        or features
End
f=bfofitness( currentswim,data2);
                // calculate Fitness function (fs)
              {
if f==1
reducedindexcount=reducedindexcount+1;
Bforeduced2(reducedindexcount)=currentswim;
End
       }
End
     }
}
```

This algorithm may be used for the automation of the BFO with error free approach .

7. References

1. Anbazhagan S, Nair AM. Geographic Information System and groundwater quality mapping in Panvel Basin, Maharashtra, India. Journal of Environmental Geology. 2000; 45(6):753–61.
2. Durgadevagi S, Annadurai R, Meenu M. Spatial and temporal mapping of groundwater quality using GIS based Water Quality Index (A Case Study of SIPCOT-Perundurai, Erode, Tamil Nadu, India). Indian Journal of Science and Technology. 2016 Jun; 9(23):1–8.
3. Rai B, Tiwari A, Dubey VS. Identification of groundwater prospective zones by using remote sensing and geoelectrical methods in Jarla and Raniganj coalfields, Dhanbad district, Jharkhand state. Journal of Earth System Sciences. 2005; 114(5):515–22.
4. Carver S. Integrating multi-criteria evaluation with Geographic Information Systems. International Journal Geographical Information Science. 1991; 5(3):321–39.
5. Das S, Behra SC, Kar A, Narendra P, Guha NS. Central Ground Water Board. Report on Hydrogeology and Groundwater Potential of Allahabad District U.P. New Delhi: CGWB; 1985.
6. Antony SA, Balakrishnan M, Gunasekaran S, Natarajan RK. A correlation study of the ground water quality in the Manali Petroleum Industrial Region in Tamil Nadu, India. Indian Journal of Science and Technology. 2008 Nov; (6):1–11.
7. Mulat S, Yohannes F, Rashid M. Soil and erosion assessment of lake Alemaya watershed, Ethiopia. Journal of Land Degradation and Development. 2007; 17(3):333–41.
8. Sudarshan MR, Jayapradha A, Lavanya, Amarnath DJ. Evaluation of groundwater quality at Oragadam –A GIS Approach. Indian Journal of Science and Technology. 2016 Apr; 9(14):1–8.
9. Castro VE, Yang J. A fast and robust general purpose clustering algorithm. Proceedings of the Fourth European Workshop on Principles of Knowledge Discovery in Databases and Data Mining (PKDD 00); Lyon, France. 2000; 1886:208–18.
10. Castro VE. Why so many clustering algorithms? SIGKDD Explorations. Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining. 2002; 4(1):65–75.
11. Bai L. Fast global k-means clustering based on local geometrical information. Information Sciences. 2013; 215:168–80.
12. Vora P. A survey on k-mean clustering and Particle Swarm Optimization. International Journal of Science and Modern Engineering. 2013; 1(3):1–3. ISSN: 2319-6386.
13. Ramakrishnan M Modified K-means algorithm for effective clustering of categorical data sets. International Journal of Computer Applications. 2014; 89(7):1–4.
14. Verma A. Performance enhancement of K-means clustering algorithms for high dimensional data sets. International Journal of Advanced Research in Computer Science and Software Engineering. 2014; 4(1):5–9.
15. Zaghoul Z. Arabic text classification based on features reduction using Artificial Neural Networks. IEEE UKSim 15th International Conference on Computer Modelling and Simulation (UKSim); USA. 2013. p. 485–90.
16. Roy D. Genetic K-mean clustering for mixed numeric and categorical datasets. International Journal of Artificial Intelligence and Applications. 2010; 1(2):1–6.
17. Alckmin D. Hybrid Genetic Algorithm applied to the clustering problem. Revisit Investigation Operational. 2012;33(2):141–51.
18. Das S, Biswas A, Dasgupta S, Abraham A. Bacterial foraging optimization algorithm: Theoretical foundations, analysis and applications, in foundations of computational intelligence global optimization. Springer. 2009; 203:23–55.
19. Liu Y, Passino KM. Biomimicry of social foraging bacteria for distributed optimization: Models, principles, and emergent behaviors. Journal of Optimization Theory and Applications. 2002; 115(3):603–28.

20. Tripathi S, Shrivastava JK. Role of constraints on process parameters in Artificial Neural Network based simulation of composting of agricultural waste. Indian Journal of Science and Technology. 2011 May; 4(5):1–5.

21. Anuradha A. Neural network approach for text classification using relevance factor as term weighing method. IJCA Journal. 2013; 68(17):37–41.