IJCTA, 9(17) 2016, pp. 8827-8835

A Novel Approach to Integration of Case Based Reasoning and Biogeography Based Optimization for Exploring Groundwater Possibility

Ravinder Kaur*, Sanjay Singla* and Sukhdeep Singh*

ABSTRACT

Groundwater Exploration is the one of the burning issue in nature and environmental science. At present time, India is leading towards major water crises. Groundwater will be the most exploited natural resource in future. Therefore, it is need of the time to have efficient and precise exploration of groundwater resources. There can be various methods to save earth from these groundwater crises like groundwater management, groundwater reuse or exploration of new resources for groundwater.

In this paper, we have proposed a novel integrated approach of Case based reasoning and biogeography based optimization for groundwater exploration. The proposed concept is structured in the manner to have the input in the form six attributes of slope, geology, landuse, lineament, soil & landform and give output in the form of low, intermediate and high possibility. In this method, case based reasoning used in the manner to retrieve the previous knowledge of use cases. Biogeography based optimization has the strategy to explore the optimum groundwater possibility in optimized manner. The overall method is evaluated using the parameters of sensitivity, specificity and accuracy.

Keywords: Biogeography Based Optimization, Case Based Reasoning, Ground Water Exploration, Optimization

I. INTRODUCTION

Water is the prime natural resources, a basic human need and a precious national asset. From existence of humans on earth it is water that guides the settlements that decide what to grow and what to sow, people need it, fight for it and above all can't survive without it. But the resources of water are decreasing with time. There is no single "magic bullet" that can resolve this rising water deficiency problem. One of the major resources of fresh water on the earth surface is ground water [1].

Groundwater is among the precious resource of nation and is permanent and reliable source of water for a majority of the world's population, especially in the arid and semi-arid regions of the world, where surface water is either scanty or strongly seasonal in occurrence.

On the earth surface, groundwater is available in very less amount. Due to unseasonal rain on earth surface, people usually depend upon groundwater [2]. Therefore, groundwater is an important commodity which we use for various purposes such as agricultural, industrial and domestic use but with the increase in population its resources are depleting and hencethe necessity to find its resources arises. Here detection for possibility of ground water at a particular region is taken as an application.

The considered proposed concept is integration of BBO and CBR method for the exploration off groundwater. Groundwater is explored for the different land cover features in the form of possibility of low,

* Computer Science Engineering, IET Bhaddal Technical Campus, City: Roopnagar, Country: India, *E-mails: rbanwait45@yahoo.com; hodcse@ietbhaddal.edu.in; Sukhdeep.cs53@ietbhaddal.edu.in*

intermediate and high results. The overall resultsare calculated using the parameters namely Specificity, Sensitivity and Accuracy.

Other sections of the paper are described as: Section II presented the considered basic concepts for proposed approach, Section III presents the considered dataset, Section IV discuses about proposed concept, Section V shows the calculated results based on the considered parameters and Section VI concludes the paper.

II. BASIC CONCEPTS

A. **Biogeography Based Optimization :** The idea of Biogeography Based Optimization was first presented by Simon in 2008 [3]. Biogeography Based Optimization is also a swarm based algorithm that is inspired by the migrating behaviour of the species as per the ecosystem. The natural process of migration of species from one island to another according to the survival circumstances is adapted for the generation of this algorithm. In BBO, the islands are considered as the solution of the problem & migration behaviour is taken as sharing of feature solutions [4]. This natural phenomenon can be used for the solution of various hard computational problems.

Biogeography can be defined as the geographical dissemination of biological species. It is the natural phenomenon that is inspired by the species to change their geographical places as per the suitable ecosystem of their habitat. In this migration process, the whole species of particular animals move from one place to another one.

The science of biogeography can be traced to the work of nineteenth century naturalist such as Wallace [5] and Charles Darwin [6]. In BBO for each individual species, a habitat with Habitat Suitability Index (HSI) is assigned. Also, Suitability Index Variable (SIV) is used to characterize the habitually of an island. For a good solution value, the HSI should be high [5].

B. **Case Based Reasoning :** Case-based reasoning is an Artificial Intelligence based approach but it works differently as compare to other AI based approaches in the manner that CBR uses previously experience based knowledge instead of solely dependent on problem domain, their description and available resources. The previous experienced based knowledge is considered as the cases for the problem solution. These cases are considered as the iterations to solve the problem. Another advantage of CBR approach is have the incremental solution for each time due to repetition of results improved which leads to overall higher efficient solution of each problem. [8].

The internal structure of CBR mechanism is categorized into components: case reasoner and the case retriever [9]. The appropriate cases in the case base can be retrieve by case retriever and further case reasoner uses the retrieved cases to find the solution of the problem [10]. This reasoning process generally involves both determining the differences between the cases retrieved and the current case, and modifying the solution to reflect these differences appropriately. The components of CBR system are as shown in figure 1.

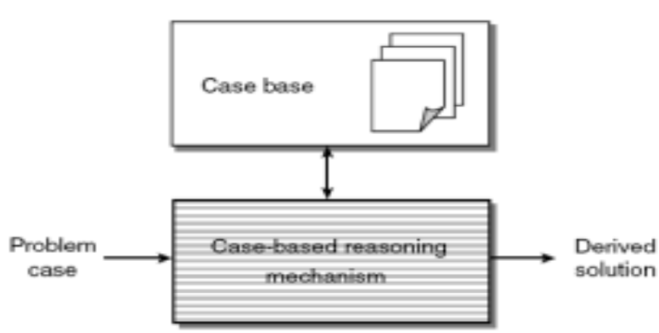
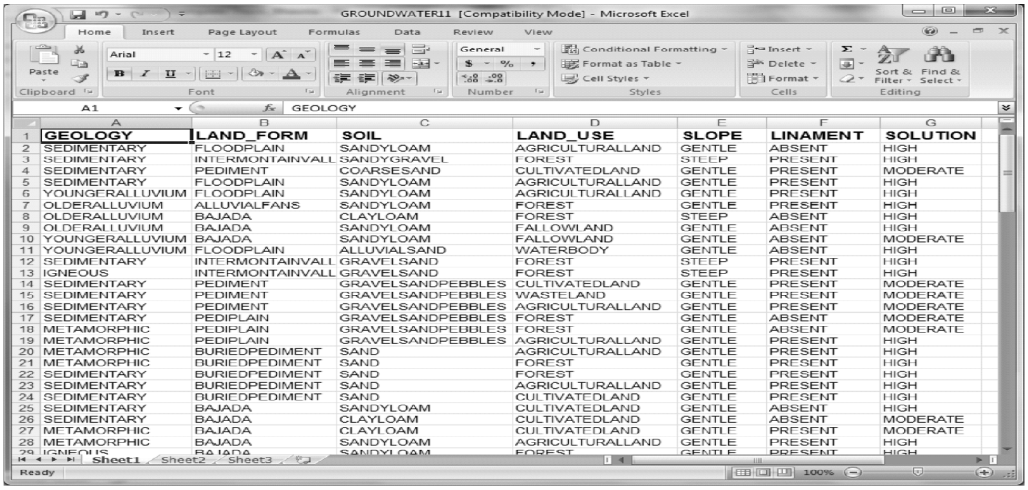


Figure 1: Case Based Reasoning

III. DATASET CONSIDERED

The proposed integrated concept is used to explore the groundwater possibility. For this,we have used the expert dataset having six attributes of slope, landuse, landform, geology, soil type and lineament. The considered dataset can be used for any location to test the possibility of groundwater.



	GEOLOGY	LAND_FORM	SOIL	LAND_USE	SLOPE	LINAMENT	SOLUTION
2	SEDIMENTARY	FLOODPLAIN	SANDYLOAM	AGRICULTURALLAND	GENTLE	ABSENT	HIGH
3	SEDIMENTARY	INTERMONTAINVALL	SANDYGRAVEL	FOREST	STEEP	PRESENT	HIGH
4	SEDIMENTARY	PEDIMENT	COARSESAND	CULTIVATEDLAND	GENTLE	PRESENT	MODERATE
5	SEDIMENTARY	FLOODPLAIN	SANDYLOAM	AGRICULTURALLAND	GENTLE	PRESENT	HIGH
6	YOUNGERALLUVIUM	FLOODPLAIN	SANDYLOAM	AGRICULTURALLAND	GENTLE	PRESENT	HIGH
7	OLDERALLUVIUM	ALLUVIALFANS	SANDYLOAM	FOREST	GENTLE	PRESENT	HIGH
8	OLDERALLUVIUM	BAJADA	CLAYLOAM	FOREST	STEEP	ABSENT	HIGH
9	OLDERALLUVIUM	BAJADA	SANDYLOAM	FALLOWLAND	GENTLE	ABSENT	HIGH
10	YOUNGERALLUVIUM	BAJADA	SANDYLOAM	FALLOWLAND	GENTLE	ABSENT	MODERATE
11	YOUNGERALLUVIUM	FLOODPLAIN	ALLUVIALSAND	WATERBODY	GENTLE	ABSENT	HIGH
12	SEDIMENTARY	INTERMONTAINVALL	GRAVELSAND	FOREST	STEEP	PRESENT	HIGH
13	IGNEOUS	INTERMONTAINVALL	GRAVELSAND	FOREST	STEEP	PRESENT	HIGH
14	SEDIMENTARY	PEDIMENT	GRAVELSANDPEBBLES	CULTIVATEDLAND	GENTLE	PRESENT	MODERATE
15	SEDIMENTARY	PEDIMENT	GRAVELSANDPEBBLES	WASTELAND	GENTLE	PRESENT	MODERATE
16	SEDIMENTARY	PEDIMENT	GRAVELSANDPEBBLES	AGRICULTURALLAND	GENTLE	PRESENT	MODERATE
17	SEDIMENTARY	PEDIPLAIN	GRAVELSANDPEBBLES	FOREST	GENTLE	ABSENT	MODERATE
18	METAMORPHIC	PEDIPLAIN	GRAVELSANDPEBBLES	FOREST	GENTLE	ABSENT	MODERATE
19	METAMORPHIC	PEDIPLAIN	GRAVELSANDPEBBLES	AGRICULTURALLAND	GENTLE	PRESENT	HIGH
20	METAMORPHIC	BURIEDPEDIMENT	SAND	AGRICULTURALLAND	GENTLE	PRESENT	HIGH
21	METAMORPHIC	BURIEDPEDIMENT	SAND	FOREST	GENTLE	PRESENT	HIGH
22	SEDIMENTARY	BURIEDPEDIMENT	SAND	FOREST	GENTLE	PRESENT	HIGH
23	SEDIMENTARY	BURIEDPEDIMENT	SAND	AGRICULTURALLAND	GENTLE	PRESENT	HIGH
24	SEDIMENTARY	BURIEDPEDIMENT	SAND	CULTIVATEDLAND	GENTLE	PRESENT	HIGH
25	SEDIMENTARY	BAJADA	SANDYLOAM	CULTIVATEDLAND	GENTLE	ABSENT	HIGH
26	SEDIMENTARY	BAJADA	CLAYLOAM	CULTIVATEDLAND	GENTLE	ABSENT	MODERATE
27	METAMORPHIC	BAJADA	CLAYLOAM	CULTIVATEDLAND	GENTLE	PRESENT	MODERATE
28	METAMORPHIC	BAJADA	SANDYLOAM	AGRICULTURALLAND	GENTLE	PRESENT	HIGH
29	IGNEOUS	BAJADA	SANDYLOAM	FOREST	GENTLE	PRESENT	HIGH

Figure 2: Considered Training Dataset

Different attributes are described with their possible features. These six attributes are further subcategorized into their respective fields as shown in table 1.

Table 1
Attributes and their Subcategories

Attributes	*Values*
Lineament	Present, Absent
Slope	Steep, Gentle
Geology	Metamorphic, Igneous, Sedimentary, Older alluvium, Younger alluvium
Land use	Wasteland, Forest, Grass, Fallowland, Swampy land, Cultivatedland, Shrubs, Buildup, Agriculturalland, Urban, Waterbody, mixed vegetation etc.
Soil	Gravelsand, Sandygravel, Coarsesand, Sand, Clayloam, Alluvialsand, Gravel Sand Pebbles, Sandyloam, Rocky etc.
Landform	DeltaicPlain, Floodplain, Pediment, Bajada, Riverterraces, Alluvialfans, Pediplain, Buriedpediment, AlluvialPlain, Intermontanevalley, Wadi, Oldmeander etc.

IV. PROPOSED CONCEPT

In this section, the integrated proposed concept of Case Based Reasoning and Biogeography Based Optimization is presented for groundwater exploration without digging the borewell. In this integrated approach, Case Based Reasoning uses previously experience based knowledge instead of solely dependent on problem domain, their description and available resources. Biogeography Based Optimization is swarm intelligence based concept best known for the global search strategy solution. Here, we have considered the integrated approach to optimize the solution upto the possible extent for groundwater possibility detection.

In this proposed concept, user query is considered as the input and groundwater possibility is determined as output. Initially, use cases are created using the expert dataset which are actually the habitat suitable condition for BBO species. BBO species try to find the best suitability conditions with quality values. Here, suitability solutions are determined by using cosine similarity function. Then BBO algorithm is applied to find the similarity of expert data cases and calculated one's.After the number of iterations, we can get the possibility results of low, moderate or higher.Complete formulation is given below:

Input: User Query.

Output: Estimation of Groundwater (Low, Moderate, High).

ALGORITHM

Step 1: Consider the training dataset for the use cases.

Step 2: Insert the user query for groundwater detection with some attribute values

Step 3: Evaluate the Feature weights for the user query data and training dataset as shown in equation (1).

$$wt = numberofoccurance.log_2\left(\frac{totalnumberoffeatures}{numberofsamplesthedataoccuredin}\right) \quad \text{...Equation (1)}$$

Step 4: Initialize the population of BBO species and set the value of max_iterationas per total case bases.

Step 5: For max_iteration{

Calculate the Cosine similarity of input query with the available case bases by considering equation (2):

$$Similarity = cos(\_) = \frac{wt_A . wt_B}{\|wt_A\|\|wt_B\|} = \frac{\sum_{i=1}^{n} wt_{Ai} * wt_{Bi}}{\sqrt{\sum_{i=1}^{n}(wt_{Ai})^2} * \sqrt{\sum_{i=1}^{n}(wt_{Bi})^2}} \quad \text{...Equation (2)}$$

Where, wt_A is the query data weight and

wt_B is the data weight of some particular case base.

5.2 Memorise the similarity for each case base.}

Step 6: Initialize the value for series = 1.

Step 7: Apply the concept of Biogeography Based Optimization.

7.1. Create habitats (population) H_1, H_2, . . . ,H_m corresponding to each land cover feature consisting of corresponding training data (m is the number of different land cover features).

7.2. Consider the original dataset as the universal habitat consisting of defined values of the species (i.e. pixels)

7.3. Compute corresponding HSI values for each habitat using Mean of standard Deviation as

$$HSI_i = \frac{\sum_{k=1}^{mb} S_{i,k}}{mb} \quad \text{...Equation (3)}$$

where, $S_{i,k}$ represents standard deviation of k^{th} band of i^{th} habitat, i is for each habitat (different features exist on the study area), varies from 1 to m., (k varies for each available subattribute, i.e., 1 to mb, where mb is the number of features available)

$$S_{i,k} = \sqrt{\frac{(\sum_{j=1}^{n_i} v_i(k,j) - M_{i,k})^2}{n_i}} \quad \text{...Equation (4)}$$

Where, n_i is the training data available for each attributes.

$v_i(k,j)$ describes the subattributes value of k^{th}attribute in j^{th} training dataset of i^{th} habitat

$M_{i,k}$ is the mean of k^{th} band in i^{th} habitat

$$M_{i,k} = \frac{\sum_{j=1}^{n_i} v_i(k,j)}{n_i} \quad \text{...Equation (5)}$$

7.4. For each species in universal habitat

Compute immigration rate » and emigration rate μ ,S_{max} and S_{min} for each habitat based on HSI;

7.4.1.For each habitat H_i

Recompute HSI values for each habitat using equations 3, 4 and 5 by combining the training data and the species data together. (i.e. the suitability of the habitat H_i is matched with a species emigrated to this habitat).

If Recomputed HSI is within the threshold value T

Migrate that species to the habitat H_i

End **If**

End **For** loop

End **For** loop

Step 8: Store the value of best solution.

Step 9: The best case is determined by the index of best solution.

Step 10: Post process results{

10.1. Apply all the available propositional logic condition

10.2. As per available condition, output will be either

{Low, Intermediate or High}}

Step 11: Set series = series + 1.

Step 12: Repeat Steps 4 to 11 until initialized series reach to Maximum Series Number.

V. RESULTS & DISCUSSION

This section determines the evaluated results for the proposed algorithm in the form of groundwater possibility to be low, intermediate or high. Also the overall comparative parameters of specificity, sensitivity and accuracy are evaluated.

(A) Results

The proposed integrated concept of BBO & CBR is implemented in MATLAB with GUI (Graphical User Interface). The considered input attributes are landuse, lineament, soil type, landform, slope and geology. Output is shown with the groundwater possibility of Low, Moderate and higher. The low probability shows groundwater possibility upto 64%. Moderate probability shows groundwater possibility from 65% to 84% and higher values are from 85% to 100%. A sample GUI is shown in figure 3.

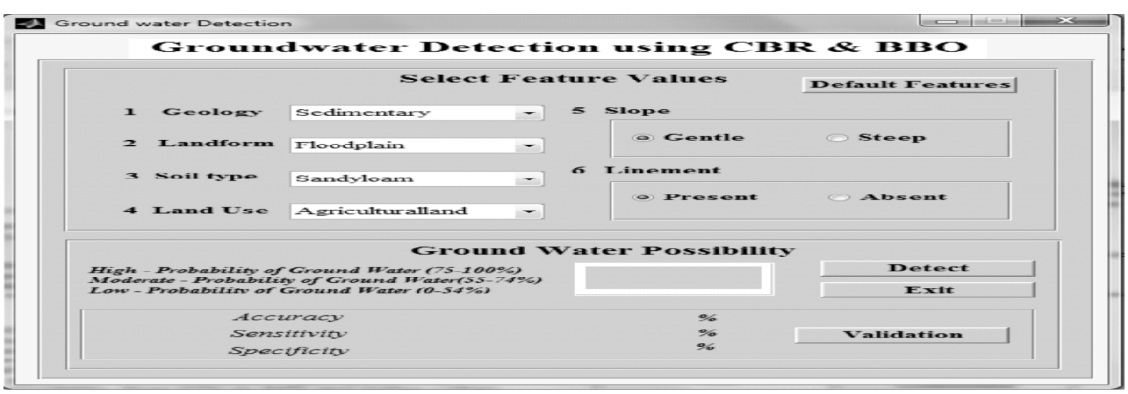


Figure 3: GUI for Groundwater Exploration

As per the shown figure 3, output is determined in low, moderate and high possible value solutions. For the output results, we have considered some test cases which as shown as below.

1). Test Case 1 : In test case 1, the considered attributes with their subcategories are shown in table 2.

Table 2
Test Case 1

Attribute	*Attribute Value*
Lineament	Absent
Slope	Steep
Land Use	Urban
Soil type	Loam
Landform	Wadi
Geology	Igneous
Groundwater possibility	?

The considered subattributes of Lineament: Absent, Slope: Steep, Landuse: Urban, Soil type: Loam, Landform: Wadi and Geology: Igneous. For these attributes, output results with performance evaluation parameters are shown in figure 4.

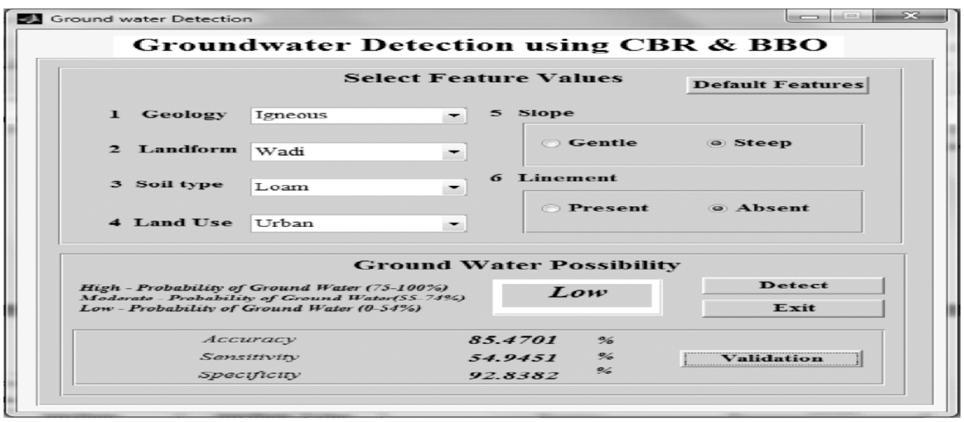


Figure 4: Test Case 1 for Groundwater Possibility

Figure 4 indicates the "**Low**" probability of groundwater possibility.

2). Test Case 2: In test case 2, the considered attributes with their subcategories are shown in table 3.

Table 3
Test Case 2

Attribute	*Attribute Value*
Lineament	Present
Slope	Gentle
Land Use	Shrubs
Soil type	Clayloam
Landform	Bajada
Geology	Olderalluvium
Groundwater possibility	?

The considered subattributes of Lineament: Present, Slope: Gentle, Landuse: Shrubs, Soil type: Clayloam, Landform: Bajada and Geology: Olderalluvium. For these attributes, output results with performance evaluation parameters are shown in figure 5.

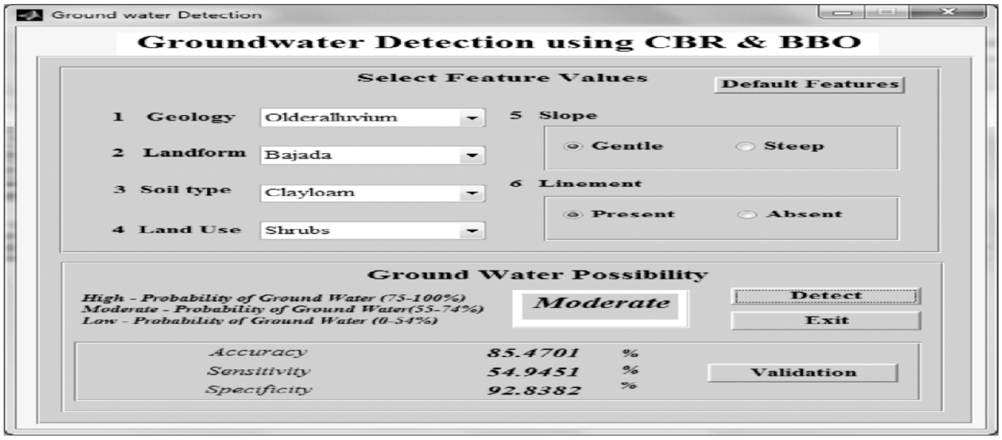


Figure 5: Test Case 2 for Groundwater possibility

Figure 5 indicates the "**Moderate**" probability of groundwater possibility.

3). Test Case 3 : In test case 3, the considered attributes with their subcategories are shown in table 4.

Table 4
Test Case 3

Attribute	*Attribute Value*
Lineament	Present
Slope	Gentle
Land Use	Forest
Soil type	Sandyloam
Landform	Pediment
Geology	Sedimentary
Groundwater possibility	?

The considered subattributes of Lineament: Present, Slope: Gentle, Landuse: Shrubs, Soil type: Clayloam, Landform: Bajada and Geology: Olderalluvium. For these attributes, output results with performance evaluation parameters are shown in figure 6.

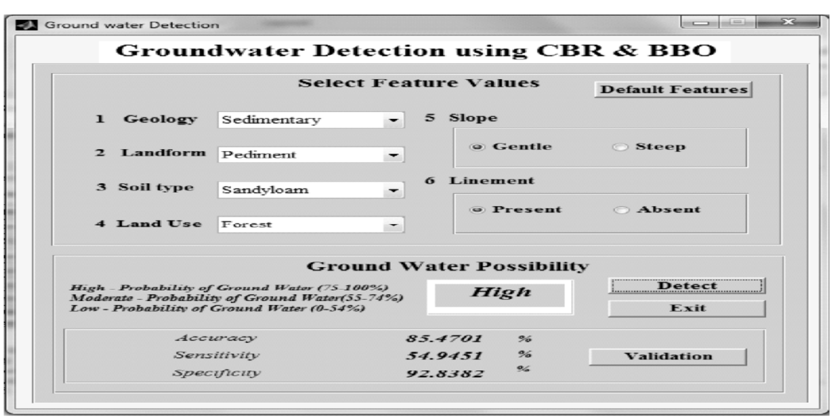


Figure 6: Test Case 3 for Groundwater possibility

Figure 6 indicates the "**High**" probability of groundwater possibility.

From the considered test cases, we can say that groundwater possibility varies in the form of Low, Moderate and High. There are many more cases for the possibility of low, moderate and high. The considered dataset cases are compared with the results of proposed algorithm and calculated sensitivity, specificity and accuracy values.

(B) Evaluation Parameters

1) *Specificity:* Itdenotes the proportion to measurethe positiveness of concept. In other words, we can say that it is the method to have the accurate value of groundwater possibility as per expert dataset. This can be calculated as:

$$Specificity = \frac{TP}{TP + FP} \qquad \text{...Equation (6)}$$

2) Sensitivity: Itdenotes the proportion to measurethe negativeness of concept. In other words, we can say that it is the method to have the inaccurate value of groundwater possibility as per expert dataset. This can be calculated as:

$$Specificity = \frac{FP}{TP + FP} \qquad \text{...Equation (7)}$$

3) Accuracy: It defines the combined values of Sensitivity and Specificity. It can be calculated as below:

$$Accuracy = \frac{\Sigma TP + \Sigma TN}{\Sigma Total\ Dataset\ Cases} \qquad \text{...Equation (8)}$$

From the output results and evaluation parameters as shown in figure 4 to figure 6, values are:

Table 6
Parametric values

Parameter	*Values (%)*
Specificity	92.8382
Sensitivity	54.9451
Accuracy	85.4701

This can also be respresented in the form of graphical representation as shown in figure 7.

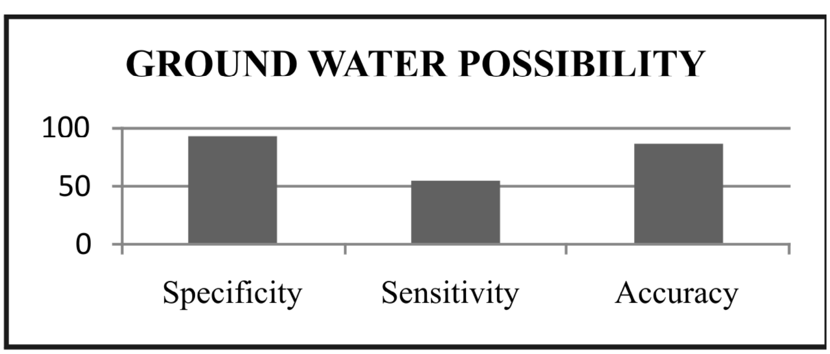


Figure 7: Ground water possibility

VI. CONCLUSIONS

Groundwater is the ubiquitous source of water due to its availability in all the seasons as compare to other water resources. But due to increasing use and more industrialization & urbanization, groundwater level decreasing continuously. So, there is the need of some efficient method to find the more groundwater resources. In this paper, we have applied the Case Based Reasoning and Biogeography Based Optimization method for the exploration of groundwater resources. Case Based Reasoning uses previously experience based knowledge instead of solely dependent on problem domain, their description and available resources. Biogeography Based Optimization is swarm intelligence based concept best known for the global search strategy solution. Here, we have considered the integrated approach to optimize the solution upto the possible extent for groundwater possibility detection. The possibility of groundwater is determined by low, moderate and high level as shown in figure 4 to 6. Also the evaluated parameters show optimized results as shown in table 6 and figure 7. So, we can conclude with possibility to determine the groundwater efficient.

REFERENCES

[1] Sincero, Arcadio Pacquiao. *Environmental Engineering: A design approach*. Vol. 1. Pearson College Division, 1996.

[2] Saraf, A. K., and P. R. Choudhury. "Integrated remote sensing and GIS for groundwater exploration and identification of artificial recharge sites."*International journal of Remote sensing* 19, no. 10 (1998): 1825-1841.

[3] Simon, D. (2008). Biogeography-based optimization. Evolutionary Computation, IEEE Transactions on, 12(6), 702-713.

[4] Ma, H. (2010). An analysis of the equilibrium of migration models for biogeography-based optimization. Information Sciences, 180(18), 3444-3464.

[5] Wallace, A. R. (2011). The geographical distribution of animals: with a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface (Vol. 1). Cambridge University Press.

[6] Darwin, Charles. "The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life." (2009).

[7] Ma, Haiping, and Dan Simon. "Blended biogeography-based optimization for constrained optimization." *Engineering Applications of Artificial Intelligence*24, no. 3 (2011): 517-525.

[8] Richter, Michael M., and Rosina O. Weber. "Case-Based Reasoning." *A Textbook* (2013).

[9] Watson, Ian, and FarhiMarir. "Case-based reasoning: A review." *The knowledge engineering review* 9, no. 04 (1994): 327-354.

[10] deMantaras, Ramon Lopez. "Case-based reasoning." In *Machine Learning and Its Applications*, pp. 127-145. Springer Berlin Heidelberg, 2001.