

Electric Profiling Based Fully Recurrent Deep Neural Learning Classification for Groundwater Quality Prediction

Raghuveer Narsing^{1,*}, K. Karthikeyan²

¹Department of Civil Engineering, Annamalai University, Chidambaram, Tamilnadu, India

²Department of Civil Engineering, Faculty of Engineering, Annamalai University, Chidambaram, Tamilnadu, India

Received December 2, 2021; Revised January 22, 2022; Accepted February 16, 2022

Cite This Paper in the following Citation Styles

(a): [1] Raghuveer Narsing, K. Karthikeyan , "Electric Profiling Based Fully Recurrent Deep Neural Learning Classification for Groundwater Quality Prediction," *Civil Engineering and Architecture*, Vol. 10, No. 2, pp. 425-437, 2022. DOI: 10.13189/cea.2022.100203.

(b): Raghuveer Narsing, K. Karthikeyan (2022). *Electric Profiling Based Fully Recurrent Deep Neural Learning Classification for Groundwater Quality Prediction*. *Civil Engineering and Architecture*, 10(2), 425-437. DOI: 10.13189/cea.2022.100203.

Copyright©2022 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Groundwater is present under Earth surface within soil pore spaces and rock formation. It is recharged via surface and typically discharged. Water pollution affects the quality of water and troubles human health, economic growth as well as social wealth. The groundwater quality identification is essential to maintain freshwater resources for sustainable development. But, the existing methods failed to improve the groundwater quality and minimize time consumption. To address these problems, an Electric Profiling Ground Water Identification based Fully Recurrent Deep Neural Learning Classification (EPGWI-FRDNLC) Method is designed to achieve efficient quality analytics by higher accuracy and minimum time consumption. In EPGWI-FRDNLC Method, electric profiling process is carried out for ground water identification. After that, a fully recurrent deep neural learning classification process is carried out for ground water quality prediction analytics. Fully recurrent deep neural learning classification process includes more than three layers for performing the ground water quality analysis. In EPGWI-FRDNLC Method Model, a lot of data were measured for input and given to the input layer. After that, input data were given to hidden layer 1. In that layer, softmax regression is used for performing the input parameter analysis like temperature, pH, turbidity, salinity, nitrates and phosphates. Then, the regression coefficient value is transferred to hidden layer 2. Tanimato similarity function is employed for identifying the similarity between

the regression coefficient value of training data and threshold value. Tanimato similarity value ranges from 0 to 1 and the results are sent to the output layer. By this way, EPGWI-FRDNLC Method improves the ground water quality prediction analytics. Experimental evaluation of EPGWI-FRDNLC Method was performed with various metrics by an amount of data.

Keywords Water Quality, Fully Recurrent Deep Neural Learning, Softmax Regression, Freshwater Resources, Tanimato Similarity

1. Introduction

Water quality is referred for chemical, physical and biological features of water with compatibility to particular usage. In order to recreate, drinking, fisheries, agriculture or industry activity, water is employed. A hybrid multivariate prediction model termed principal component analysis (PCA) with relevance vector machine (RVM) was introduced in [1] for seawater quality estimation. However, prediction accuracy was not minimized by PCA-RVM. A statistical model was designed in [2] depending on IoT for water quality estimation by various metrics namely pH, conductivity, and so on. However, time consumption is not reduced with statistical model.

Water Quality Monitoring (WQM) was commercial in [3] to monitor quality of drinking quality with IoT technology. But, computational cost was not reduced by WQM technique. A sensor-based water quality monitoring scheme was introduced in [4] with microcontroller for system processing as well as communication scheme. But, water quality monitoring accuracy was not improved by the designed scheme.

Bayesian Maximum Entropy (BME) was introduced in [5] to identify Water Quality Monitoring Stations (WQMS) to attain maximum information by minimal number of monitoring stations. Mask-RCNN image segmentation method was introduced in [6] to attain fish and illustrate trajectory image of fish at a particular time. Trajectory image datasets were partitioned within normal and abnormal water quality. But, error rate is not minimized with Mask-RCNN image segmentation method.

An integrated method was designed in [7] to identify the sampling points. In order to find exact locations, Multi criteria assessment technique was used. But, the accuracy was not at the required level by integrated method. A scenario-based numerous-objective optimization scheme was introduced in [8] for spatio-temporal optimal plan. However, time complexity was not minimized by the designed scheme.

A multivariate statistics method was introduced in [9] to attain historical temporal water quality variation because of water circuit modification. But, water quality identification accuracy was not improved by multivariate statistics method.

River mixing length scheme (RML) is designed in [10] with sampling points. A non-point source potential pollution score (NPPS) is introduced to categorize sampling point for river system. However, the complexity level was not minimized by RML method.

The issues identified from the above literature are high computational cost, higher computational complexity, lesser water quality monitoring accuracy, higher error rate, higher water quality monitoring time consumption, etc. In order to address these issues, an Electric Profiling Ground Water Identification based Fully Recurrent Deep Neural Learning Classification (EPGWI-FRDNLC) Method is introduced.

The main contributions of the EPGWI-FRDNLC Method are:

- The aim of the EPGWI-FRDNLC Method is used to achieve effective quality analytics by higher accuracy and minimal time consumption. In EPGWI-FRDNLC Method, an electric profiling process is carried out for ground water identification.
- Fully recurrent deep neural learning classification process is carried out for ground water quality prediction analytics. In EPGWI-FRDNLC Method Model, number of data were measured and given to input layer. After that, input data were given to hidden layer 1.

- In EPGWI-FRDNLC Method, Softmax Regression Classification (SRC) is used at the hidden layer 1 for performing the input parameter analysis. The novelty of the Softmax regression is used to measure the probability with every class for classifying the data points. Based on the exponential value, the classifier predicts the particular class into different classes. The output of hidden layer 1 is transmitted to hidden layer 2.
- In EPGWI-FRDNLC Method, the novelty of the Tanimato similarity function is employed at the hidden layer 2 for finding the similarity between the regression coefficient value of training data and threshold value. Tanimato similarity value ranges from 0 to 1 and outcomes are sent to output layer. In this manner, EPGWI-FRDNLC Method improves the ground water quality prediction analytics.

The article is structured as follows. Section 2 reviews related work of water quality monitoring using different methods. Section 3 explains proposed EPGWI-FRDNLC Method by neat diagram. Section 4 describes experimental setup with dataset. Section 5 describes different performance results and parameters. Section 6 explains the conclusion of this article.

2. Related Works

A smart water quality monitoring system was introduced in [11] for in-pipe water quality with power efficient manner by IoT. Designed system alert remote user was divergence of water quality. However, error rate is not reduced with smart water quality monitoring system.

A cost effective IoT based smart water quality monitoring scheme was introduced in [12] to monitor quality metrics. Designed scheme was experienced by water samples and metrics were sent to server. A low cost real-time water quality monitoring system was introduced in [13] for the rivers, lakes, coastal areas as well as water bodies. However, computational complexity is not reduced by designed scheme.

LSTM network model is designed in [14] with parameters like pondus hydrogenii (PH), dissolved oxygen (DO), and so on. However, water quality is not exactly identified by LSTM model. Water quality monitoring scheme was introduced in [15] to examine fish behavior for monitoring water pollution. But, computational cost was not reduced by a designed system.

Hybrid water quality monitoring device fault diagnosis scheme was designed in [16] with multiclass support vector machines (MSVM) and rule-based decision trees (RBDT). But, complexity is reduced and accuracy was not improved.

3-D dynamic scheme depending on Lagrange technique was introduced in [17] to forecast the propulsive performance with continuous diving and surfacing motions.

However, error rate is not reduced with 3-D dynamic model.

A multiple fish tracking algorithm was introduced in [18] for red snapper. Otsu adaptive segmentation extracted fish targets by background elimination technique. However, multiple fish tracking algorithm failed to reduce the time complexity.

A low-cost unmanned surface vehicle (USV) was introduced in [19] for water quality analysis. The water environmental mobile observer (WeMo) developed for considering chemical and physical metrics. But, computational cost was not reduced by low-cost USV. A water quality monitoring system was introduced in [20] for advising users by water quality. Designed method computed the physiochemical water quality parameters. But, the accuracy level was not improved by water quality monitoring system.

3. Proposed Methodology

Water covers 71% of the earth's surface and it is significant natural resource. 2.5% of earth water was fresh and suitable for consumption. Water quality estimation on watershed scale needs water pollution and detection of major pollution aspects. Therefore, water quality prediction needs to be carried out through groundwater identification and quality analysis. Many researchers carried out their research on groundwater identification and water quality prediction. But, accuracy is not enhanced and time consumption is not minimized. In order to address these problems, an Electric Profiling Ground Water Identification based Fully Recurrent Deep Neural Learning Classification (EPGWI-FRDNLC) Method is introduced. EPGWI-FRDNLC is employed for enhancing performance water quality prediction in terms of accuracy and time consumption. Architectural diagram of EPGWI-FRDNLC Method is illustrated in figure 1.

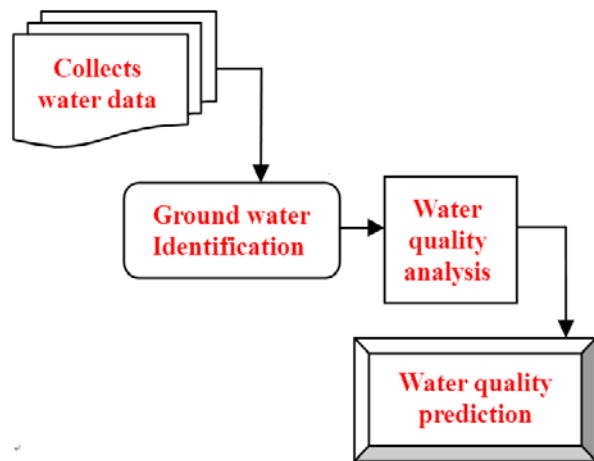


Figure 1. Flow process of EPGWI-FRDNLC Method

Fig. 1 explains the architectural diagram of

EPGWI-FRDNLC method for water quality prediction by higher accuracy and minimal time consumption. Water quality dataset was taken as input for examining the groundwater identification and quality prediction. The groundwater identification is performed using Electrical Profiling process. After identifying the groundwater, the quality level is computed using recurrent deep learning classifier based on the physical, chemical and biological parameters like temperature, pH, nitrates, and so on. The two processes of the EPGWI-FRDNLC method are described briefly in the subsections.

3.1. Electrical Profiling Groundwater Identification

In EPGWI-FRDNLC Method, electric profiling process is carried out for ground water identification. Electric profiling establishes the resistivity variations on the horizontal scale for performing the ground water identification. Groundwater is an essential part of water resources. Groundwater was described by subsurface water which fills in soil and breaks rock formation. Groundwater was considered by another water supply for every living thing. In order to detect groundwater zone in subsurface layer, issue addressed with engineers. Geophysical was a utilization of physics which investigated earth through considering capacity on earth surface. Geophysical is a mainly appropriate tool for examination of groundwater. Electrical Profiling Resistivity Method (EPRM) was the part of geophysical technique involved in any groundwater exploration. EPRM is used to identify the thickness of layered media for mapping the geological environment. EPRM is used efficiently for groundwater detection because of simplicity as well as non-destructive development for generating subsurface imaging.

Electrical Profiling (EP) is carried out within Segen River by two profile lines. Profile line one is on dam axis and profile line 2 is on upper dam axis. It is aligned in North-East and South-West direction. EP is carried out with the profile line one on dam axis by profile length of 960m. Profile line two is carried out on upper dam axis by length of 1025m by sampling intervals of 5m, 10m, 20m and 30m. EP is used for finding relative position and geological structures and litho logic contacts for dam construction. EP for profile line one on 30m depth is classified with heterogeneous resistivity. It proves the higher resistivity ranging from 35–120 Ω -m of profile line. Resistivity on middle of profile gets reduced to 35–10 Ω -m with exception of profile line resistivity at 300–400m from the Teklehaimanot church. Electrical profiling demonstrates resistivity value is higher 35 Ω -m. Contour map of profile 1 illustrates graben-like boundary. EP contour map plotted for profile 2 on 30m depth was categorized with highest heterogeneous resistivity.

Electrical resistivity profiling is a two-dimensional resistivity scheme where images represent slice of earth. In order to interpret contradictory geological or

hydrogeological conditions, Electrical property contrasts were employed. Color contouring is employed for detecting situation. Two-dimensional resistivity profiling is employed for enhancing borehole data or locating new borehole with spatially extending information away from hole. George Simon Ohm describes that an electric current ‘ I ’ in wire is proportional to voltage ‘ V ’. It is given by,

$$V = IR \tag{1}$$

From (1), ‘ R ’ represents the conductor resistance. The unit of resistance is measured in Ω . George Simon Ω is formulated as resistance is proportional to length ‘ L ’ and inversely proportional to cross sectional area ‘ A ’ of conductor. It is computed by,

$$R = \rho \frac{L}{A} \tag{2}$$

From (2), ‘ ρ ’ denotes resistivity of material. Resistivity is a physical property of material that expresses their ability to oppose flow of charge. It is given by

$$\rho = \frac{AV}{LI} \tag{3}$$

From (3), resistivity was employed for shallow subsurface and groundwater exploration. In order to calculate resistivity, four electrodes were applied. One electrode is employed for penetrating current into ground. In order to calculate potential difference among hemispherical equi-potential surfaces, additional pair is utilized where they intersect ground surface. It was a conventional method of determining resistivity within uniform half space. The algorithmic process of electrical profiling resistivity groundwater identification is shown below,

Algorithm 1: Electrical Profiling Resistivity Groundwater Identification

Input: Dataset, Station
Output: Identify the groundwater

- Begin**
1. for each station,
 2. Determine the voltage level ‘ V ’
 3. Calculate the apparent resistivity ‘ ρ ’
 4. Identify groundwater in the location of a particular station
- 5. End for**
End

Algorithm 1 illustrates the process of the electrical profiling resistivity groundwater identification. Let us consider that location in a particular station to identify the groundwater level. The optimal distance is computed based on the voltage level and apparent resistivity. Depending on the calculations, the groundwater around the river delta region is identified.

3.2. Fully Recurrent Deep Neural Learning Classification Based Predictive Analytics

Deep Learning is the division of machine learning that uses calculation for procedure data and performs analysis process. Deep learning uses the layer of calculation for procedure the data, understand human discourse with observe objects. In EPGWI-FRDNLC Method, recurrent deep neural learning classification method comprised four layers, namely input layer, two hidden layers, output layer for performing efficient water quality prediction analytics. The structure of fully recurrent deep neural learning is illustrated in figure 2.

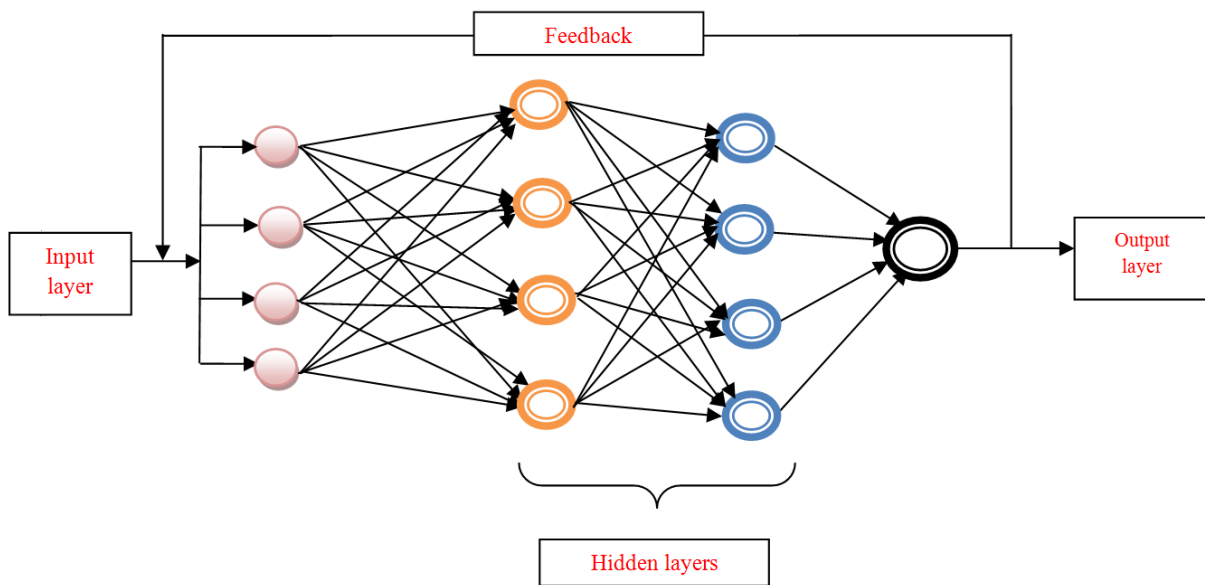


Figure 2. Construction of Fully Recurrent Deep Neural learning Classifier

Fig 2 explains construction of Fully Recurrent Deep Neural learning Classifier for performing efficient water quality prediction analytics with better accuracy with minimal time. As described in figure 2, the neural network comprised different layers like one input and output layer with two hidden layers. Each layer included the neurons like activating nodes that are fully linked from one layer to another layer. Input layer gathers amount of data at the time 't' represented by 'I_a(t)'. After that, the neurons were connected from one layer to another layer through dynamic weights. It is formulated as,

$$I_a(t) = \sum_{i=1}^n d_i * w_i + b \tag{4}$$

From (4), 'I_a(t)' denotes the neuron activity on input layer. Input is sent to hidden layer 1. In that layer, EPGWI-FRDNLC Method uses the Softmax Regression Classification (SRC) analysis for performing ground water identification and quality prediction analytics. SRC analysis is a generalization of logistic regression that handles multiple classification problems. Softmax regression is carried out with the data features like temperature, pH, turbidity, salinity, nitrates and phosphates. The SRC classifier output is mathematically attained as,

$$SRC = \frac{\exp(C_j)}{\sum_{i=1}^n \exp(\vartheta_{d_i})}$$

Where $j = 1, 2, 3 \dots m$ (5)

From (5), 'SRC' denotes the softmax regression classifier result for each input data. The output of SRC classifier is used as the probability distribution over 'm' dissimilar possible outcomes and the predicted probability of 'jth' 'jth' class with extracted features (ϑ). After that, the predicted probability is computed as,

$$P(\gamma = j|\vartheta_{d_i}) = \frac{\exp(c_j)}{\sum_{i=1}^n \exp(\vartheta_{d_i})} \quad j = 1, 2, 3 \dots m \tag{6}$$

From (6), the SRC classifier decides the higher exponential values. SRC classifier is used to determine the probability of each class to categorize input data points. For each input data, the number of output classes is obtained as follows,

$$W(\vartheta_{d_i}) = \begin{bmatrix} P(\gamma = 1|\vartheta_{d_i}) \\ P(\gamma = 2|\vartheta_{d_i}) \\ \vdots \\ P(\gamma = m|\vartheta_{d_i}) \end{bmatrix} = \frac{1}{\sum_{i=1}^n \exp(\vartheta_{d_i})} \begin{bmatrix} \exp(c_1) \\ \exp(c_2) \\ \vdots \\ \exp(c_m) \end{bmatrix} \tag{7}$$

From (7), SRC classifier forecasts the specific class among 'm' number of classes according to exponential value. The regression coefficient value is obtained at the layer and sent to hidden layer 1. Next, hidden layer 1 result

is transferred to hidden layer 2. Tanimoto similarity function is employed to identify the similarity between the regression coefficient value of training data and threshold value for identifying the water quality. Tanimoto similarity between the two data points are determined as follows,

$$\delta = \frac{n * \sum p_{n_1} * td}{\sum p_{n_1}^2 + \sum td^2 - \sum p_{n_1} * td} \tag{8}$$

From (8), 'δ' denotes the similarity coefficient. 'n' denotes the number of data points. '∑ p_{n₁} * td' symbolizes the sum of the product of paired scores, 'p_{n₁}, td' are the two training data and threshold value 'td'. '∑ p_{n₁}²' represents the sum of squared score of the training data, ∑ p_{n₂}² represents a sum of a squared score of the threshold data. It is denoted by,

$$H(t) = [\sum_{i=1}^n d_i * w_i] + [w_{ih} * H_{t-1}] \tag{9}$$

From (9), 'H(t)' represents hidden layer on time period 't' and \square_{t-1} 'H_{t-1}' denotes output of previous hidden layer, 'd_i' denotes the input data, 'w_{□□}' 'w_{ih}' symbolizes the weight among input and hidden layer. Tanimoto similarity coefficient provides the output results from '0' to '1'. Lastly, the error rate is determined at the output of hidden layer to increase the classification accuracy and reduce the error rate. The error is determined as the squared difference between the actual and predicted results expressed as,

$$Error = (Actual\ result - Obtained\ result)^2 \tag{10}$$

From (10), the error rate is determined. When the network not accurately classifies the particular data, the proposed algorithm adjusted the weights and identified the minimum error. It is obtained as,

$$\arg\min t_E \tag{11}$$

From (11), 'arg min' symbolizes the argument of minimum function. Similarity values are transformed within the output layer of deep neural network. Similarity values are analyzed with the threshold value. The output layer is formulated as,

$$O(t) = \{H(t) * w_{ho}\} \tag{12}$$

From (12), 'O(t)' represents the output layer result. 'w_{ho}' represents the weight allocated between the hidden layer and output layer. Data is classified into five types (i.e., drinking, irrigation, industries, and domestic consumption and water bodies) for ground water quality prediction by better accuracy and minimal time consumption. After the classification, the data correctly predicts the water quality level with higher prediction accuracy and minimum error rate. The algorithmic process of the proposed EPGWI-FRDNLC Method is illustrated as,

Algorithm 2: EPGWI-FRDNLC Method**Input:** Number of data**Output:** Increase ground water quality prediction**Begin**

1. Collect a number of data at input layer
2. The input layer transmits data to the hidden layer 1
3. **For** each data
4. Hidden layer 1 uses the softmax regression to analyze and identify the water quality level
5. Hidden layer 2 uses the tanimoto similarity function
6. Compute similarity between training data and threshold at the hidden layer 2
7. Calculate the error value
8. Update initial weights
9. Find *arg min error*
10. Output layer displays the result

11. End for**End**

Algorithm 2 explains the algorithmic process of EPGWI-FRDNLC Method for ground water quality prediction. EPGWI-FRDNLC Method considers data for Deep Recurrent Neural Learning Classifier. Next, hidden layer 1 analyzes the data by using softmax regression classifier. After that, the result is transferred to hidden layer 2. Tanimoto similarity function is used to determine the similarity between input data and threshold data. When the error is minimal, the result is sent to the output layer. Similarly, output of hidden layer 2 is feedback and specified as input for hidden layer 1. Finally, output layer displays the final result for ground water quality prediction.

4. Experimental Settings

Experimental analysis of the proposed EPGWI-FRDNLC Method and two existing methods, namely PCA-RVM [1] and statistical model [2], is implemented using MATLAB simulator. In order to conduct the experiment, two datasets are employed, namely Indian water quality dataset (D1) and water samples gathered from the river bed of ANDHRA PRADESH state (D2). With the help of these two datasets, the water quality level gets examined for irrigation and drinking purposes. Indian water quality dataset is collected from

<https://www.kaggle.com/anbarivan/indian-water-quality-d-ata>. The dataset has 1991 instances and 12 attributes (i.e. columns). The dataset includes historical data on pollution levels and temperature across different states in India. For experimental consideration, number of data considered as input ranging from 100 to 1000. Correspondingly, the water quality parameters gathered from the river bed of ANDHRA PRADESH state and conduct experiment with

the data ranging from 10 to 100. The 12 attributes descriptions are listed below:

Table 1. Attribute description

S. No	Attributes	Description
1	Station code	Station code of every listed place
2	LOCATIONS	Address of the place where the data acquired from
3	STATE	Name of the state where the values are measured
4	temp	Temperature in °C
5	D.O. (mg/l)	Dissolved oxygen in (mg/l)
6	PH	Potential of Hydrogen
7	CONDUCTIVITY (µmhos/cm)	Water can conduct an electrical current
8	B.O.D. (mg/l)	Biochemical Oxygen Demand ...
9	NITRATENAN N+ NITRITENANN (mg/l)	Chemical element
10	FECAL COLIFORM (MPN/100ml)	fecal coliform bacteria
11	TOTAL COLIFORM (MPN/100ml)Mean	Total coliforms include bacteria that are found in water
12	year	Year of data collected

5. Performance Analysis and Discussion

In the analysis and discussion section, the experimental results of the proposed EPGWI-FRDNLC Method and existing methods namely PCA-RVM [1] and statistical model [2] are explained. The comparison of the three methods is performed with help of table and graph. The analysis of different methods is carried out in the subsection. The performance of EPGWI-FRDNLC Method is determined by three different parameters like prediction accuracy, error rate and prediction time.

5.1. Impact of Prediction Accuracy

It is described as proportion of number of data was correctly predicted to the number of data points collected as dataset. The prediction accuracy is formulated as,

$$\text{Prediction Accuracy} = \left[\frac{\text{Number of correctly predicted data points}}{\text{Total number of data points}} \right] * 100 \quad (13)$$

From (13), prediction accuracy is measured. It is calculated by percentage (%).

Sample Calculation for Dataset 1:

- **Existing PCA-RVM:** Total number of data points is 100 and the number of data points precisely predicted is 86. Thus, the prediction accuracy is determined as follows,

$$\text{Prediction Accuracy} = \left[\frac{86}{100} \right] * 100 = 86\%$$

- **Existing Statistical Model:** Total number of data points is 100 and the number of data points precisely predicted is 89. Thus, the prediction accuracy is determined as follows,

$$\text{Prediction Accuracy} = \left[\frac{89}{100} \right] * 100 = 89\%$$

- **Proposed EPGWI-FRDNLC Method:** Total number of data points is 100 and the number of data points precisely predicted is 95. Thus, the prediction accuracy is determined as follows,

$$\text{Prediction Accuracy} = \left[\frac{95}{100} \right] * 100 = 95\%$$

Table 2(a) and (b) describes prediction accuracy by number of data using two different datasets. The prediction accuracy is computed based on number of data considered as input dataset. From Table 2, prediction accuracy results

denote that the EPGWI-FRDNLC Method attains better results than the existing methods PCA-RVM [1] and Statistical model [2]. In the dataset 1, 100 input data are considered as input. At the first iteration, 95 data are correctly predicted using EPGWI-FRDNLC Method through performing the classification with the prediction accuracy of 92%. By using two existing methods, PCA-RVM [1] and Statistical model [2], the number of data correctly predicted is 86 and 89 and their prediction accuracy attained is 86% and 89% correspondingly. Similarly, the prediction accuracy results are attained for remaining nine iterations. Correspondingly, chemical properties of water data from the ANDHRA PRADESH state were determined. Ten results of prediction accuracy are determined for every method. The average of ten results symbolizes that water quality prediction accuracy of EPGWI-FRDNLC Method is higher than the other two existing methods. Figure 3 (a) (b) describes the graphical illustration of prediction accuracy for two different datasets.

Table 2(a). Prediction accuracy using dataset (D1)

Number of data	Prediction accuracy (%)		
	EPGWI-FRDNLC Method	PCA-RVM	Statistical Model
100	95	86	89
200	98	88	90
300	95	86	90
400	96	85	89
500	96	86	90
600	97	85	88
700	98	87	91
800	98	88	93
900	98	89	92
1000	98	88	92

Table 2(b). Prediction accuracy using dataset (D2)

Number of data	Prediction accuracy (%)		
	EPGWI-FRDNLC Method	PCA-RVM	Statistical Model
10	90	70	80
20	95	85	90
30	93	83	90
40	95	88	90
50	98	88	92
60	97	87	92
70	99	86	93
80	98	85	90
90	98	87	91
100	97	86	92

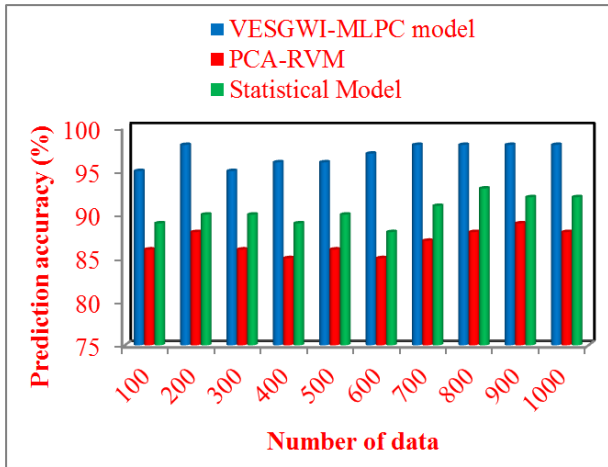


Figure 3(a). Analysis of Prediction accuracy using dataset 1

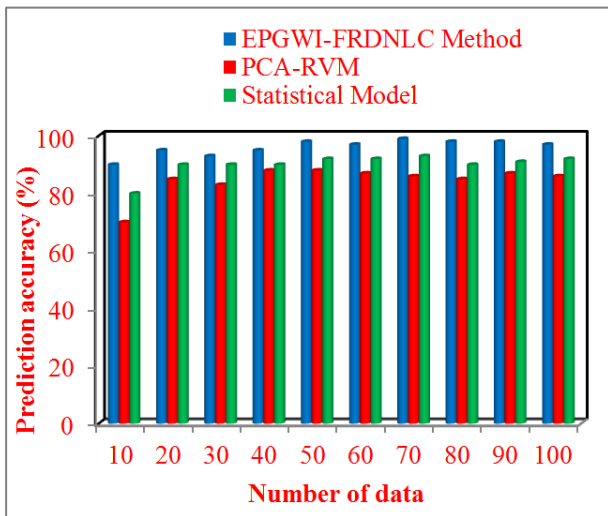


Figure 3(b). Analysis of Prediction accuracy using dataset 2

Figure 3 (a) and (b) explains the water quality prediction accuracy results by different number of data. Number of input data was considered on horizontal axis and prediction accuracy of three different methods are considered on vertical axis in terms of percentage. As described in the graph, there are three colours of cylinders namely blue, red, and green denotes the prediction accuracy of three methods, namely EPGWI-FRDNLN Method, PCA-RVM [1] and Statistical model [2], respectively. The graphical illustration of EPGWI-FRDNLN achieves two existing methods. This is because of applying an electric profiling

process for ground water identification. After that, a fully recurrent deep neural learning classification process is carried out for ground water quality prediction analytics. Softmax regression is used in neural learning process for performing the input parameter analysis like temperature, pH, turbidity, salinity, nitrates and phosphates. Tanimoto similarity function identifies the similarity between regression coefficient value of training data and threshold value. The average value represents that prediction accuracy of the EPGWI-FRDNLN Method is improved by 12% and 7% when compared with [1] and [2] in dataset 1. In addition, the prediction accuracy is increased by 14% compared with [1] and 7% compared with [2] in dataset 2.

5.2. Impact of Error rate

Error rate is described by proportion of number of data incorrectly predicted to number of data gathered from the input dataset. Consequently, the error rate is formulated as,

$$Error\ rate = \left[\frac{Number\ of\ data\ incorrectly\ predicted}{Number\ of\ data} \right] * 100 \quad (14)$$

From (14), error rate is computed. It is determined by percentage (%).

Sample Calculation for Dataset 1:

- **Existing PCA-RVM:** Number of data points is 100 and the number of data points incorrectly predicted is 14. Thus, the error rate is measured as follows,

$$Error\ rate = \left[\frac{14}{100} \right] * 100 = 14\%$$

- **Existing Statistical Model:** Number of data points is 100 and the number of data points incorrectly predicted is 11. Thus, the error rate is measured as follows,

$$Error\ rate = \left[\frac{11}{100} \right] * 100 = 11\%$$

- **Proposed EPGWI-FRDNLN Method:** Number of data points is 100 and the number of data points incorrectly predicted is 5. Thus, the error rate is measured as follows,

$$Error\ rate = \left[\frac{5}{100} \right] * 100 = 5\%$$

Table 3(a). Error rate using dataset (D1)

Number of data	Error Rate (%)		
	EPGWI-FRDNLC Method	PCA-RVM	Statistical Model
100	5	14	11
200	2	12	10
300	5	14	10
400	4	15	11
500	4	14	10
600	3	15	12
700	2	13	9
800	2	12	7
900	2	11	8
1000	2	12	8

Table 3(b). Error rate using dataset (D2)

Number of data	Error Rate (%)		
	EPGWI-FRDNLC Method	PCA-RVM	Statistical Model
100	10	30	20
200	5	15	10
300	7	17	10
400	5	12	10
500	2	12	8
600	3	13	8
700	1	14	7
800	2	15	10
900	2	13	9
1000	3	14	8

Table 3(a) and (b) explains the performance result of error rate by the number of data with two different datasets. The error rate is computed depending on the number of data taken as input dataset. From table 3, error rate results denote that the EPGWI-FRDNLC Method attains better results than the existing methods PCA-RVM [1] and process-based watershed model-ANN [2]. In the dataset 1, 100 data are considered. In the first iteration, 5 data are incorrectly predicted through classification using EPGWI-FRDNLC Method and error rate obtained is 5%. With two existing methods, namely PCA-RVM [1] and Statistical model [2], the number of data incorrectly predicted are 14 and 11 and their error rate is 14% and 11%, respectively. Likewise, the error rate results are attained for three different methods. The average of ten results denotes that error rate of EPGWI-FRDNLC Method is

less than the other two existing methods. Figure 4 (a) (b) illustrates the graphical illustration of error rate for two different datasets.

Figure 4 (a) (b) explains the experimental results of error rate depending on number of data points. As described in the above graph, the proposed EPGWI-FRDNLC Method obtains lesser error rate for two datasets. This is due to the application of an electric profiling process for ground water identification. Fully recurrent deep neural learning classification process is carried out for performing the ground water quality prediction analytics. Softmax regression performs parameter analysis such as temperature, pH, turbidity, salinity, nitrates and phosphates. Tanimato similarity function computes the similarity between regression coefficient value of training data and threshold value.

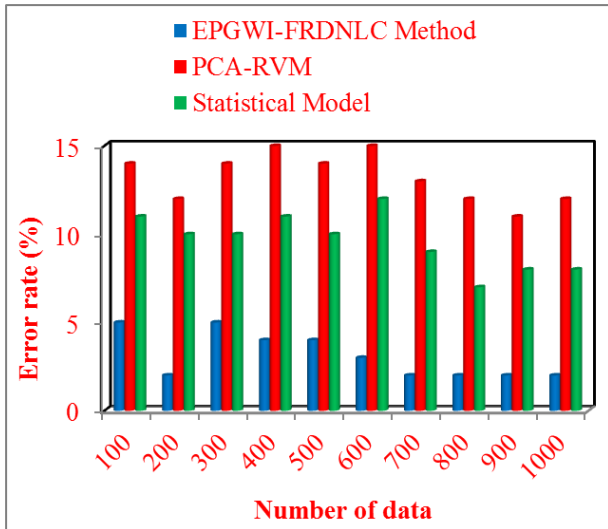


Figure 4(a). Analysis of error rate using dataset 1

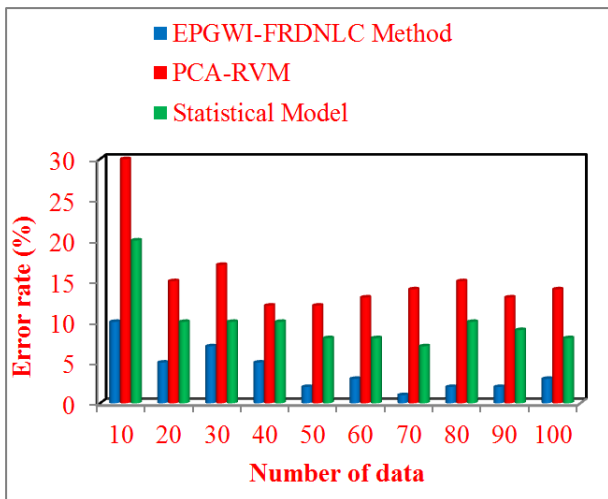


Figure 4(b). Analysis of error rate using dataset 2

The average error rate value of the EPGWI-FRDNLC Method is reduced by 77% and 68% when compared to existing PCA-RVM [1] and Statistical model [2], respectively in dataset 1. In addition, the error rate is reduced by 75% when compared to PCA-RVM [1] and 62% when compared to Statistical model [2] in dataset 2.

5.3. Impact of Prediction Time

It is the number of time taken to forecast the water quality by using the classification process. The water quality prediction time is computed as,

$$Prediction\ time =$$

$$n * time\ consumed\ for\ predicting\ one\ data \quad (15)$$

From (15), prediction time is computed. ‘n’ denotes the number of data collected from input dataset. The prediction time is computed in terms of milliseconds (ms).

Sample Calculation for Dataset 1:

- **Existing PCA-RVM:** Number of data points is 500 and the time utilized to predict the water quality data points is 0.1296 ms. Thus, the processing time is computed as,

$$PT = 100 * 0.2ms = 20\ ms$$

- **Existing Statistical Model:** Number of data points is 500 and the time utilized to predict the water quality data points is 0.1296 ms. Thus, the processing time is computed as,

$$PT = 100 * 0.16\ s = 16\ ms$$

- **Proposed EPGWI-FRDNLC Method:** Number of data points is 500 and the time utilized to predict the water quality data points is 0.1296 ms. Thus, the processing time is computed as,

$$PT = 100 * 0.1ms = 10\ ms$$

Table 4(a). Prediction time using dataset (D1)

Number of data	Prediction Time (ms)		
	EPGWI-FRDNLC Method	PCA-RVM	Statistical Model
100	10	20	16
200	12	24	18
300	15	27	21
400	17	30	23
500	19	33	25
600	21	35	27
700	22	39	28
800	25	42	30
900	27	45	31
1000	28	48	33

Table 4(b). Prediction time using dataset (D2)

Number of data	Prediction Time (ms)		
	EPGWI-FRDNLC Method	PCA-RVM	Statistical Model
100	6	11	8
200	7	12	9
300	9	14	11
400	10	15	12
500	11	16	13
600	12	17	14
700	13	18	14
800	13	19	15
900	14	21	16
1000	15	22	17

The graphical analysis of water quality prediction time versus number of data from two different datasets and three methods namely EPGWI-FRDNLC Method, PCA-RVM [1], Statistical model [2] are described in table 4(a) (b). The observed table values denote that VESGWI-MLPC model reduced the groundwater quality prediction time consumption. The experiment is carried out for 100 data from dataset 1, the time consumption of the proposed EPGWI-FRDNLC Method to predict the water quality is 10ms whereas the time consumption of prediction using existing PCA-RVM [1], Statistical model [2] are 20ms and 16ms, respectively. Similarly, additional results are achieved with different input data. Figure 5 (a) (b) describes the graphical illustration of prediction time for two datasets.

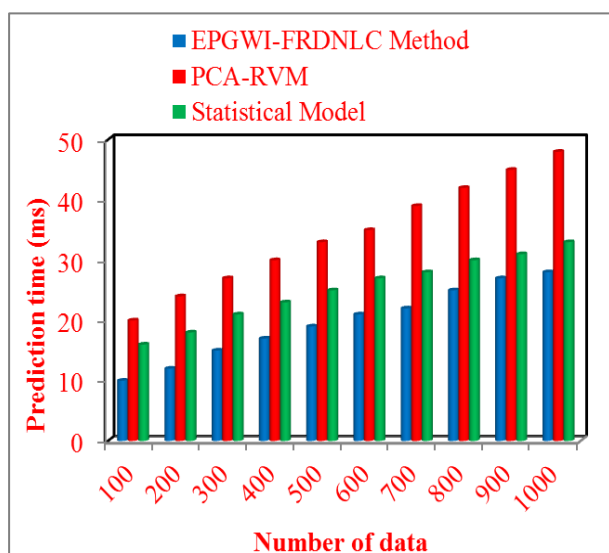
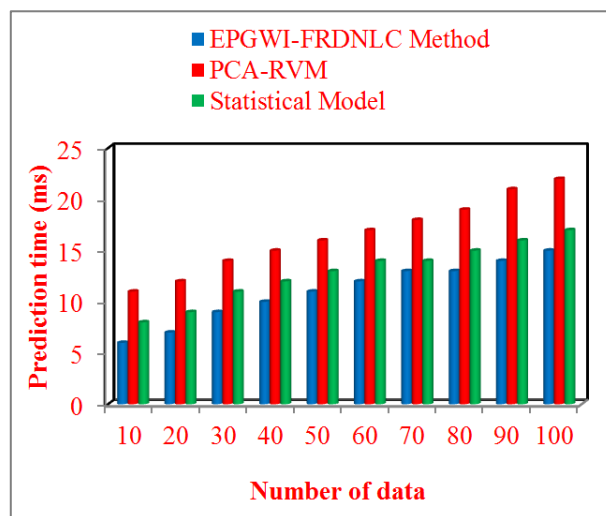
**Figure 5(a).** Analysis of prediction time using dataset 1**Figure 5(b).** Analysis of prediction time using dataset 2

Figure 5 (a) (b) describe the graphical representation of water quality prediction time by various number of data from two different datasets. The number of data is taken on the horizontal axis and different prediction time performance results are taken on the vertical axis. However, quality prediction time gets reduced for two different datasets using the EPGWI-FRDNLC Method than the other two existing methods. This is because of using electric profiling process and fully recurrent deep neural learning classification process for ground water identification and ground water quality prediction analytics. Softmax regression performs the parameter analysis. Tanimoto similarity function determines the similarity between the regression coefficient value of training data and threshold value. The average of ten results represents that overall water quality prediction time is reduced by 44% when compared to [1] and 24% when compared to [2], respectively. Likewise in dataset 2, the average of ten results indicates that the overall quality prediction time gets minimized by 34% and 16% when compared to existing methods.

6. Conclusions

This paper presents a novel deep learning method called EPGWI-FRDNLC Method for groundwater quality identification. The related work recommends designing an appropriate classification method for increasing the performance of water quality prediction with lower time and enhanced accuracy. The goal of the EPGWI-FRDNLC Method is achieved to perform efficient quality prediction analytics with the implementation of electric profiling process and fully recurrent deep neural learning classification process. The EPGWI-FRDNLC method enhances the number of data points are correctly categorized as compared to existing works. The EPGWI-FRDNLC method reduces the amount of time consumed for classifying water quality into dissimilar classes as compared to conventional works. In addition, the proposed EPGWI-FRDNLC method lessens the number of images wrongly classified when compared to existing works. The proposed EPGWI-FRDNLC method is implemented in two datasets. The effectiveness of the EPGWI-FRDNLC method is measured in terms of accuracy, prediction time, and error rate as compared to different conventional works of [1], and [2]. The results verify that the EPGWI-FRDNLC Method improved prediction accuracy by 10%, reduced the error rate by 73%, and minimized prediction time by 34% compared to the state-of-the-art works using the dataset 1. Similarly, the EPGWI-FRDNLC Method using dataset 2 enhances the prediction accuracy by 11% minimizes the error rate and prediction time by 69% and 25% when compared to conventional methods.

REFERENCES

- [1] Xiaochuan Sun., Xin Wang., Di Cai., Zhigang Li., Yuanyuan Gao., Xusheng Wang, "Multivariate Seawater Quality Prediction Based on PCA-RVM Supported by Edge Computing Towards Smart Ocean", *IEEE Access*, vol. 8, pp. 54506 – 54513, 2020.
- [2] Prasad M. Pujar, Harish H. Kenchannavar, Raviraj M. Kulkarni., Umakant P. Kulkarni, "Real-time water quality monitoring through Internet of Things and ANOVA-based analysis: a case study on river Krishna", *Applied Water Science* Vol. 10, pp. 1-15., 2020.
- [3] Sathish Pasika., Sai Teja Gandla, "Smart water quality monitoring system with cost-effective using IoT", *Heliyon*, Elsevier, Vol. 6, pp. 1-15., 2020.
- [4] Mohammad Salah Uddin Chowdury., Talha BinEmran, Subhasish Ghosh, Abhijit Pathak, Mohd. Manjur Alam, Nurul Absar, Karl Andersson, Mohammad Shahadat Hossain, "IoT Based Real-time River Water Quality Monitoring System", *Procedia Computer Science*, Elsevier, Vol. 155, pp. 161-168., 2019.
- [5] Reza Salman, Mohammad Reza Nikoo, Shahab Aldin Shojaezadeh, Pouyan Hatami Bahman Beiglou, Mojtaba Sadegh, Jan Franklin Adamowski., Nasrin Alamdari, "A novel Bayesian maximum entropy-based approach for optimal design of water quality monitoring networks in rivers", *Journal of Hydrology*, Vol. 603, pp. 1-15, 2021.
- [6] Cheng Shuhong, Zhang Shijun., Zhang Dianfan, "Water quality monitoring method based on feedback self correcting dense connected convolution network", *Neurocomputing*, Elsevier, Vol. 349, pp. 301-313, 2019.
- [7] Hossein Alilou, Alireza Moghaddam Nia, Mohsen Mohseni Saravi, Ali Salajegheh, Dawei Han, Bahram Bakhtiari Enayat, "A novel approach for selecting sampling points locations to river water quality monitoring in data-scarce regions", *Journal of Hydrology*, Elsevier, Vol. 573, pp. 109-122, 2019.
- [8] Shokoufeh Pourshahabi, Gholamreza Rakhshandehroo, Nasser Talebbeydokhti, Mohammad Reza Nikoo., Fariborz Masoumi, "Handling uncertainty in optimal design of reservoir water quality monitoring systems", *Environmental Pollution*, Elsevier, Vol. 266, pp. 1-15, 2020.
- [9] Thi Minh Khanh Le., Mikko Makela., Nora Schreithofer., Olli Dahl., "A multivariate approach for evaluation and monitoring of water quality in mining and minerals processing industry", *Minerals Engineering*, Elsevier, Vol. 157, pp. 1-15, 2020.
- [10] Hossein Alilou, Alireza Moghaddam Nia, Hamidreza Keshtkar, Dawei Han and Michaela Bray, "A cost-effective, efficient framework to determine water quality monitoring network locations", *Science of The Total Environment*, Elsevier, Vol. 624, pp. 283-293, 2018.
- [11] S. Geetha., S., Gouthami., "Internet of things enabled real time water quality monitoring system", *Smart Water*, Springer, vol. 2, pp.1-19, 2016.
- [12] Varsha Lakshmikantha, Anjitha Hiriyannagowda, Akshay Manjunath, Aruna Patted, Jagadeesh Basavaiah., Audre Arlene Anthony, "IoT based Smart Water Quality Monitoring System", *Global Transitions Proceedings*, Elsevier, Vol. 2, pp. 1-15, 2021.
- [13] Alexander T. Demetillo, Michelle V. Japitana., Evelyn B. Taboada "A system for monitoring water quality in a large aquatic area using wireless sensor network technology", *Sustainable Environment Research*, Vol. 29, pp. 1-24, 2019.
- [14] Chuhan Qim, Shuo Huang., Xiaofei Wang "Monitoring Water Quality Parameters of Taihu Lake Based on Remote Sensing Images and LSTM-RNN", *IEEE Access*, Vol. 8, pp. 188068 – 188081, 2020.
- [15] Fei Yuan, Yifan Huang, Xin Chen., En Cheng, "A Biological Sensor System Using Computer Vision for Water Quality Monitoring", *IEEE Access*, Vol. 6, pp. 61535 – 61546, 2018.
- [16] Shuangyin Liu, Longqin Xu, Qiucheng Li, Xuehua Zhao., Daoliang Li, "Fault Diagnosis of Water Quality Monitoring Devices Based on Multiclass Support Vector Machines and Rule-Based Decision Trees", *IEEE Access*, Vol. 6, pp. 22184 - 22195, 2018.
- [17] Zhengxing Wu, Jincun Liu, Junzhi Yu and Hao Fang, "Development of a Novel Robotic Dolphin., Its Application to Water Quality Monitoring", *IEEE/ASME Transactions*

on Mechatronics, Vol. 22, pp. 2130-2140, 2017.

- [18] Xiaoqiang Zhao., Sheng Yan., Qiang Gao, "An Algorithm for Tracking Multiple Fish Based on Biological Water Quality Monitoring", IEEE Access, Vol. 7, pp. 15018 – 15026, 2019.
- [19] Dario Madeo, Alessandro Pozzebon, Chiara Mocenni., Duccio Bertoni, "A Low-Cost Unmanned Surface Vehicle for Pervasive Water Quality Monitoring", IEEE Transactions on Instrumentation and Measurement, Vol. 69, pp. 1433 – 1444. 2020
- [20] Niel Andre Cloete, Reza Malekian., Lakshmi Nair, "Design of Smart Sensors for Real-Time Water Quality Monitoring", IEEE Access, Vol. 4, pp. 3975-3990, 2016