Advanced Soft Computing Methods for Groundwater Level Prediction

Alok Shrivastav

M.Tech Scholar, Department of Electronics and Communication Engineering, Ambalika Institute of Management and Technology, Lucknow, India.

Pankaj Prajapati

Assistant Professor, Department of Electronics and Communication Engineering, Ambalika Institute of Management and Technology, Lucknow, India.

Abstract—

Population expansion and pollution are depleting groundwater in growing nations like India. Monitoring groundwater levels helps manage its supplies. Groundwater estimate is key to understanding groundwater resources. Ground water modelling is non-linear and can't be solved with standard methods. Soft computing technologies like Fuzzy Logic, Genetic Algorithm, and Artificial Neural Network are becoming more significant in hydrological domain concerns. Fuzzy logic helps with dataset imprecision and ambiguity. Inspired by human learning, an artificial neural network may learn from examples and adjust weights. Genetic algorithm is inspired by nature. This paper investigates FL, ANN, and FPSO developments in groundwater level prediction. Four models are considered, using different combinations of groundwater recharge and discharge as inputs and groundwater level as an output. ANN trains, tests, and validates these models on ground water data sets to select the best model for predicting groundwater level. Models generated with FL. FL works best with two inputs and ANN with more.

Fuzzy interval optimization in FL is tricky. Hit-or-miss has been employed till now. This is why mathematicians are looking for a way to maximise fuzzy interval length. GA was designed to improve fuzzy intervals. In this method, binary GA changes the fuzzy interval length and Wang and Mendel creates the rule basis. This technique is used by three groundwater models to forecast level. The fuzzy GA method for estimating groundwater levels outperforms the FL method. This fuzzy technique predicts groundwater levels better than other methods.

Fuzzy sets are a challenge in FL. To create fuzzy sets without specialised knowledge, a computational method is devised. Methodology uses central tendency ideas. This thesis covers the design problem of finding the right amount of fuzzy sets. The suggested method locates intervals and fuzzy sets for fuzzy time series forecasting. Chennai's reservoir rains are modelled using a central tendencies-based fuzzy approach and constitute a benchmark challenge for fuzzy time series.

The anticipated values are compared to other methods' results to prove its superiority. The proposed computational technique produced encouraging results for benchmark data.

Keywords— *Ground water level, Artificial Intelligence, Fuzzy Sets, GA, ANN etc.*

I. INTRODUCTION

Groundwater is a natural resource that is absolutely necessary. In the southern region of India, groundwater is the primary source of supply for drinking water as well as industrial uses in the mining and agricultural industries [1]. The dynamic character of groundwater systems means that they are always adapting to changes in land use, climate (including temperature and precipitation), and the amount of groundwater that is withdrawn [2]. It is essential to the management of groundwater resources that one has a solid understanding of how each of these different characteristics affects the aquifer's ability to recharge, the resulting groundwater levels, and the consequent groundwater outflow. Therefore, an accurate prediction of groundwater levels is an essential step towards accomplishing sustainable management and utilization of groundwater resources.

When groundwater travels via the subsurface pathways inside the soil and the unsaturated zone, and then continues on to the saturated zone of the aquifers, the quality of the water in the ground undergoes a rapid change. It is essential to have a deeper comprehension of the inherent qualities of mineral waters, particularly groundwaters, as well as the process through which these waters acquire their features. It is necessary to have a better grasp, for the sake of groundwater management, of the governing processes and, whenever it is possible to do so, of the natural, geologically regulated baseline chemistry. This understanding is required. If one is planning on determining the effect that contaminants have on groundwaters, this is of the utmost importance [3-4]. It is necessary to have a solid understanding of the natural baseline quality if one is going to have any hope of comprehending the patterns of pollution and the effects they have on the aquifer. When it comes to defining pollution, the foundations are laid once the baseline features have been thoroughly characterised, with geological and geochemical abnormalities taken into consideration.

Simulations of groundwater flow systems have traditionally been carried out using models that are based on physical processes. The need for huge amounts of data is one of the obstacles presented by the utilization of these models [3].

In order to calibrate the model, it is necessary to define the boundary conditions and estimate the values of the other parameters. Because obtaining the information can be difficult and time consuming, it contributes to the overall cost and complexity of the modeling process [4].

Machine learning models are an appealing alternative to process-based models because they have the potential to require less effort and potentially produce useful results with fewer data samples. This makes machine learning models an appealing alternative. Machine learning models are able to recognize patterns by directly learning knowledge from the data they are fed using past data as input. These models are able to represent the trends and time-variant behaviour of hydrological systems without the in-depth understanding of the underlying physical properties that groundwater flow models require in order to function [5].

The use of artificial neural networks (ANN) for the prediction of groundwater systems has become widespread. In point of fact, numerous ANN approaches have been utilized ever since the 1990s in an effort to increase the accuracy of water level forecast [6]. Analysis of additional regression methods, such as random forest regression (RF) [7], support vector regression (SVR) [8], gradient boosting regression (GB) [9], and decision tree regression (DT) [10], has received less focus than it deserves. Therefore, the purpose of this study is to train and investigate the performance of these various regression models in predicting groundwater level variations for four boreholes located in the North West Province, and to compare their accuracy with that of a simple feed forward neural network (FFNN).

As the primary subject of this investigation, the Grootfontein aquifer was chosen since it is one of the aquifers in South Africa that is used the most frequently. Groundwater levels for four boreholes were used as model outputs, while temporal data such as monthly precipitation, temperature, and natural groundwater discharge rates from a groundwater-fed spring were used as model inputs. The monitored boreholes are only used for monitoring purposes; they are not pumped in any way. The groundwater abstraction rates were not included as model inputs since there is a lack of data on abstraction rates from neighboring boreholes. Despite the fact that regional abstraction is expected to have an effect on the measured groundwater levels, this data is extremely limited.

In South India, there is a limited amount of work in the field of forecasting groundwater levels using machine learning models. This is particularly the case in the state of Tamil Nadu. The use of machine learning models for the purpose of groundwater prediction is most common in the American continent, followed by Europe and Asia. The originality of this work is therefore centered on the application of the models that have been stated above to this specific dataset.

The primary objective of this study was to develop a method that was driven by data in order to simulate the levels of groundwater found in the Grootfontein aquifer in South India. In greater detail, this research article makes the following contributions: (a) a ranking of the qualities according to their mutual information (MI); (b) a reference for model selection; and (c) a predictive model to anticipate groundwater levels in South India.

The metrics of root mean squared error (RMSE) and r-squared (R^2) were utilized in the process of model evaluation. In addition to this, factor analysis was carried out in order to provide a quantitative understanding of the connections between the various features and the groundwater levels. According to the findings, the variables that contributed the most to the ability to forecast groundwater levels were discharge rates from the groundwater, followed by temperature and precipitation. The GB algorithm's ability to achieve R^2 scores for all four boreholes was cited as providing the most accurate results. For the vast majority of the boreholes, the ANN received the lowest possible performance scores. The general methodology of the investigation is depicted in Figure 1.

In South India, there is a limited amount of work in the field of forecasting groundwater levels using machine learning models. This is particularly the case in the state of Tamil Nadu. The

use of machine learning models for the purpose of groundwater prediction is most common in the American continent, followed by Europe and Asia. The originality of this work is therefore centered on the application of the models that have been stated above to this specific dataset. The primary objective of this study was to develop a method that was driven by data in order to simulate the levels of groundwater found in the Grootfontein aquifer in South India. In greater detail, this research article makes the following contributions: (a) a ranking of the qualities according to their mutual information (MI); (b) a reference for model selection; and (c) a predictive model to anticipate groundwater levels in South India.

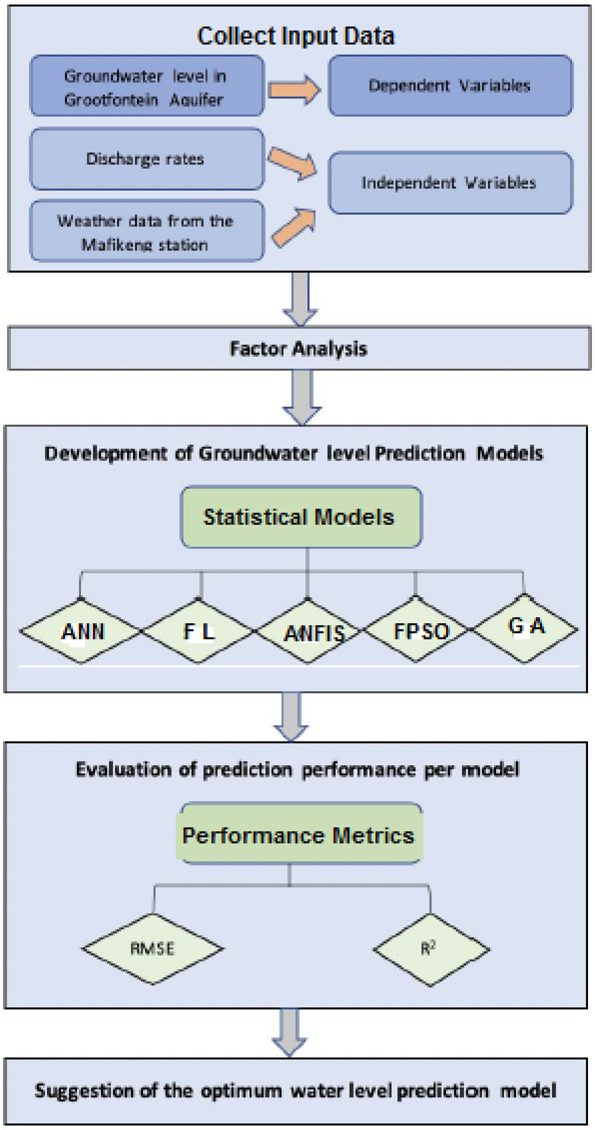

Figure 1: Flow Chart from Data Selection to final predictive model is selection

II. Literature Review

In most cases, the circumstances of actual life are marked by a high degree of imprecision, uncertainly, and vagueness. According to Zadeh [11], fuzzy logic (FL) is an idea that originates from the concept of fuzziness (vagueness). Instead of a random variable or a stochastic process being the cause of imprecision in fuzziness, the source of imprecision is a class (or classes). In order to construct the associated fuzzy set, fuzzyness works by quantitatively associating a degree of membership (ranging from 0 to 1) with each individual element of a set. The idea was developed further by Zadeh [20], who presented CW (computation with words). One of the most significant distinctions that can be made between perceptions and measurements is that, on the one hand, perceptions are hazy, while measurements are clear. When dealing with perceptions, it is vital to use a logical structure that is fuzzy rather than precise in order to handle the situation effectively. Words are how humans convey their experiences and thoughts. Because of this, manipulating perceptions becomes as simple as doing computations with words when they are first described in words (CW). Words or, more generally speaking, propositions derived from a natural language serve as the basis for computation in CW. When it comes to solving problems that occur in the real world, there is a lot to be gained by taking advantage of the tolerance for imprecision, uncertainty, and incomplete truth. This is the major impetus behind the development of both the computational theory of perceptions and the approach of computing with words.

A modelling strategy known as fuzzy rule based modelling or qualitative modelling, as defined by Sugeno et al. [12], is one in which the behaviour of the system is described using natural language. The fuzzy modelling approach is portrayed as a method to construct a system model by making use of a descriptive language that is founded on fuzzy logic and contains fuzzy predicates. Through the application of the linguistic approximation method, a qualitative model can be produced from a fuzzy model. The identification of the structure of a fuzzy model was accomplished by the application of a method known as fuzzy clustering. The linguistic approximation method has to be improved as the next step in the qualitative modelling process. In a more general sense, this signifies that scholars have made a significant breakthrough, and that computation can now be done using words rather than numbers. FL is made up of if-then rules, the clarity of whose formulation provides insights into the way in which a system behaves physically. Bardossy [13] provided two fuzzy models for the filtration process in hydrology. The first model defined rules based on the moisture content of the layers that were near to each other, and the second model defined rules based on the depth of the wetting front using the filtration rules. The second fuzzy rule-based model was retrieved from the training set that was acquired by numerically solving Richards' equation. The amount of moisture in the strata that were nearby served as the basis for the rules. The advantage of using these methods was that, in compared to traditional models, they required a smaller number of parameters and ran more quickly.

The reservoir operation modelling that was used was developed by Panigrahi [14]. The operation of single-purpose reservoirs was modelled using a fuzzy rule-based system that was built. The model is based on the logic of '*if then*' statements. Creating membership functions for the inflow, storage, demand, and release are some of the phases involved in the creation of the model. Other steps include the formulation of fuzzy rules, implication, and finally defuzzification. The Kotralai River in the Tiruvallur district of Tamil Nadu State, India, served as an example for the methodology through a case study that was conducted there. The release was the consequence of the premises of using reservoir storage, inflow, and demand as the basis for the calculation. The knowledge base required for the formulation of the fuzzy rules is provided by simulated reservoir operation coupled with a policy for maintaining a steady state that is generated from a stochastic dynamic programming model. As an alternative to the expert information that is often accessible with experienced reservoir managers, the stochastic model was utilised in this scenario.

Gharde et al. [15] established ANN and FL models to forecast the runoff and sediment yield for catchment of Kal river, India using 21 years (1991 to 2011) of rainfall and other hydrological data (evaporation, temperature, and streamflow lag by one and two days) and 7 years of data for sediment yield modelling. The models used 21 years of rainfall and other hydrological data to predict runoff and sediment yield. As the number of input vectors was increased, the performance of the ANN model got better. During both the development and validation stages, the fuzzy logic model had a performance with a R value that was greater than 0.95. When compared to the ANN model, it was discovered that the FL model performed admirably in terms of its ability to estimate runoff and sediment output for the Kal river.

The backpropogation algorithm was the catalyst for ANN's rise to prominence, which followed shortly after its discovery. Back-propagation is a learning process that was described by Rumelhart and colleagues [16] for networks of neurons. Repeated adjustments to the weights are made in accordance with the technique in order to reduce the amount of difference that exists between the network's actual output vector and the output vector that is desired for the network. As a result of the weight adjustments, the internal "hidden" units that were not a part of the input or output have come to represent key aspects of the task domain, and the interactions of these units have captured the regularities that were present in the task. This has occurred as a consequence of the fact that the weight adjustments were made. Back-propagation is distinguished from earlier, more straightforward approaches such as perceptron networks by its capacity to generate meaningful new features.

The use of fuzzy time series as a method for addressing uncertainties in data is highly effective. It was developed during the previous two decades with the goal of improving the accuracy of predicting. The enrollments of the University of Alabama were predicted by Song and Chissom (1993) using fuzzy time series as an application of fuzzy time series in educational research [17]. As a result of this, a fuzzy time series model was constructed with the use of historical data. A comprehensive method was suggested, which incorporates the following steps: fuzzy-matching the historical data; creating a fuzzy time series model; calculating and evaluating the results of the model; and so on. The reliability of the fuzzy time series model was examined so that the forecasting model could be rated appropriately. In addition, Song and Chissom presented first order temporal variant fuzzy time series models by making use of max- min com- position operation. The authors presented a strategy for the development of a temporal variant model. The neural networks defuzzification methodology had the lowest error rate out of the three distinct defuzzification methods that were used. It was discovered that simpler models generated better outcomes when used for forecasting [18].

III. STUDY AREA AND DATA COLLECTION

S. Sethuram drew the conclusion from his research that the city of Chennai's water management faces certain significant challenges, such as inadequate levels of water security. On the basis of his findings, he came to the conclusion that the water supply is erratic, unreliable, and insufficient. He attributed this to difficulties encountered in gaining access to particular regions. Due to the fact that the primary project focus of policies is on supply enhancement, demand management is severely weak. Even if there is rainwater harvesting systems in place, the demand management programmes that are now in place are not sufficient to support Chennai's expanding population. Therefore, management through conservation and efficiency requires additional enhancement and dynamic action. He suggested that Chennai needs an improved control of project management and planning because there have been several supply implementations, out of which two major failures, recorded during his data collection. While a few interview responses and survey results were contradictory to the notion that there is a water security issue that hinders water supply management in Chennai, he suggested that Chennai needs an improved control of project management and planning because there are several supply implementations.

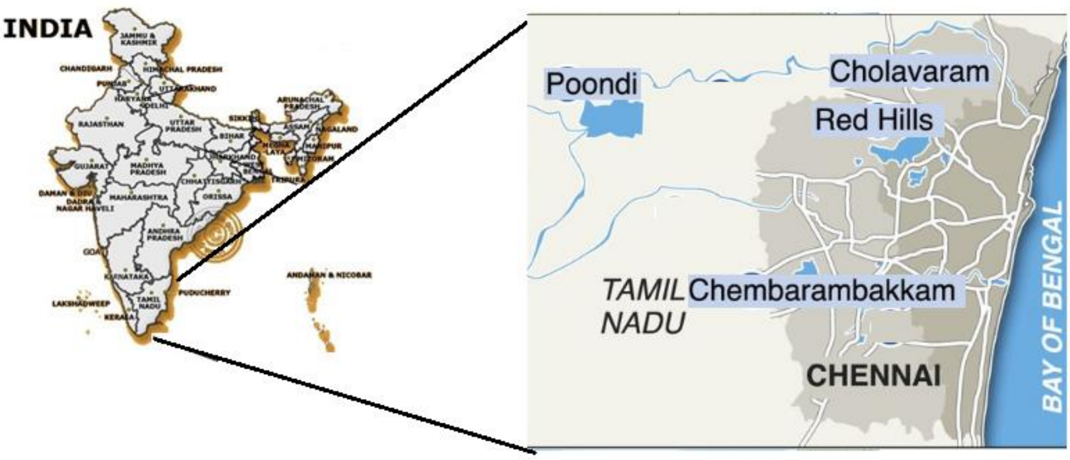

Figure 2: Location of reservoirs in Chennai (TN)

On the southeast coast of India and in the north-eastern region of Tamil Nadu is where you'll find Chennai. Its coordinates are 13.04°N 80.17°E. It is situated on what is known as the Eastern Coastal Plains, which is a very level coastal plain. The highest point in the city is 60 metres (200 feet) above sea level, while the average height of the city is 6 metres (20 feet) (200 ft). On the road, Chennai is located 2,184 kilometres (1,357 miles) to the south of Delhi, 1,337 kilometres (831 miles) to the southeast of Mumbai, and 1,679 kilometres (1,043 miles) to the southwest of Kolkata.

After the data on the annual groundwater level in the Chennai district have been gathered for the period of time from 2004 to 2020, the data are analyzed by conducting a comprehensive study of the groundwater system and filing an application for it. In created models, groundwater recharge and groundwater discharge, along with their respective time lags, function as input parameters, and groundwater level is projected to function as an output. In order to make a prediction that has a certain level of accuracy, the statistical characteristics that are used to evaluate the fitness of the observed and anticipated data are taken into consideration. In order to anticipate the level of groundwater for the next one time step, the performance of each model was evaluated, and the best model was selected based on the results. In the end, all of the models are compared in order to determine which one provides the most efficient monitoring of the groundwater level management system.

IV. PREDICTION OF GROUNDWATER LEVEL THROUGH SOFT COMPUTING TECHNIQUES

When it comes to modeling the level of groundwater, FL is particularly useful for handling ambiguity and uncertainty, whereas ANN is an effective instrument for developing data-driven models. It has been discovered that FL is helpful for calculating the amount of groundwater that is being refilled based on easily quantifiable characteristics such as rainfall and temperature [19]. ANNs are powerful modeling tools that may be used to simulate a wide variety of non-linear hydrological phenomena [58], including rainfall and runoff]. Water level projections for the Kotralai River (TN) were carried out by Bardossy et al. [20] using FL and ANN. A comparison of the simulations carried out by ANN, MFIS, and ANFIS demonstrates that ANFIS is the most useful instrument for predicting the level of groundwater [21]. The purpose of this research was to investigate the fundamental question of which method of soft computing, fuzzy logic or artificial neural networks (ANN), is superior for predicting the level of groundwater.

Analyses are carried out with the use of physical models in order to guarantee the availability of groundwater that is both high in quality and quantity. It is inevitable that the inherent complexity and data uncertainty of groundwater systems will restrict the simulation accuracy of physical-based modeling. In this study, therefore, soft computing approaches, fuzzy logic, and artificial neural networks are utilized to forecast groundwater level. The dataset that was used in the research came from the Chennai district, which is located in the Tiruvallur district of the Tamil Nadu state. The Kotralai River is located in this area. The groundwater level (W_t) for the current year is used as the output variable in the development of four models, with various combinations of the groundwater recharge (R_t) for the current year, the groundwater discharge (D_t) for the current year, the groundwater recharge for the previous year (R_{t-1}), the groundwater discharge for the previous year (D_{t-1}), and the groundwater level for the previous year (W_{t-1}) serving as the input variables [37].

Model 1: In Model 1, groundwater recharge, which is a combination of various variables such as precipitation, areal recharge, losing stream reaches and soil property, and ground-water discharge, which includes pumping extractions, temperature, dew point, and wind speed are considered as input, which require thorough investigations due to the complexity and non-linear nature of the model, while groundwater level is considered to be the output.

The complexity and non-linear nature of the model make it necessary to conduct such investigations.

$$W(t) = f\{R(t), D(t)\} \qquad (1)$$

Model 2: The groundwater recharge, groundwater discharge of the current year, and groundwater level of the previous year are the input parameters for Model 2. The groundwater level of the current year is the output parameter for this model.

$$W(t) = f\{R(t), D(t), W(t-1)\} \qquad (2)$$

Model 3: The groundwater recharge, groundwater discharge for the current year, and groundwater recharge for the previous year are the input parameters for Model 3. The output is the level of the groundwater for the current year.

$$W(t) = f\{R(t), D(t), R(t-1)\} \qquad (3)$$

Model 4: Input parameters for Model 4 include groundwater recharge, groundwater discharge for the current year, groundwater recharge and groundwater discharge for the previous year, and groundwater recharge and groundwater discharge for the year before that. The output is the level of the groundwater for the current year [38].

$$W(t) = f\{R(t), D(t), R(t-1), D(t-1)\}. \qquad (4)$$

Model 5: The groundwater recharge, groundwater discharge of the present year, and groundwater recharge of the previous year, groundwater discharge of the previous year, and groundwater level of the previous year are the input parameters for Model 5. The groundwater level of the previous year is the output.

$$W(t) = f\{R(t), D(t), R(t-1), D(t-1), W(t-1)\} \quad (5)$$

Development of Fuzzy Models: The process of mapping particular inputs to a predetermined set of outputs is referred to as FIS. There are two different kinds of FIS: M-FIS and Sugeno FIS. The M-FIS software is used to develop each and every fundamental FL model. The operation of M-FIS involves fuzzying up the inputs, building a rule base, and then defuzzifying the results in order to achieve crisp output.

Fuzzification: Due to the fact that the data for this study are in qualitative form and not fuzzy form, fuzzification was utilized so that inputs and outputs could be converted into the fuzzy form. The W_t is alienated in seven fuzzy subsets (EL, VL, L, M, H, VH, EH), that D_t is alienated in six fuzzy subsets (VL, L, M, H, VH, EH), and that R_t is alienated in five fuzzy subsets (VL, L, M, H). In this particular investigation, a triangle membership function (MF) was utilized. The membership grades are determined with the assistance of the data and under the guidance of the experts [22].

Rule base: A collection of IF-THEN rules is what the fuzzy rulebase is made out of. Rules are derived from the numerical data that has been observed over the course of the previous years and describes the level of groundwater in terms of both groundwater recharge and groundwater discharge. One such formulation of a rule could read as follows: "If R_t is EH and D_t is EL, then W_t is V H."

Defuzzification: Defuzzification is the process of obtaining a quantifiable outcome, such as a final crisp value or number that incorporates all of the characteristics of several antecedents. This can be done by obtaining a measurable result. The centroid, mean of maxima, and bisector approaches are the three most frequent types of defuzzification techniques. The centroid approach is employed as a defuzzification technique in this instance.

Development of ANN Models: There are a total of five models created for one and two hidden layers. In the case of models with two layers (hidden), the quantity of neurons is decided by a process of trial and error. The MATLAB programme known as **nntool** is used for both the training and testing of these ANN models. In this investigation, the ANN model was trained using back propagation in conjunction with the Bayesian regulation backpropagation approach. The sigmoid transfer function, also known as 'tansig,' the log sigmoid transfer function, also known as 'logsig,' and the pure linear transfer function, also known as 'purelin,' have been selected to serve as transfer functions for the hidden layer as well as the output layer [23].

Distributed processing systems make extensive use of artificial neural networks (ANNs), which are useful forms of computer intelligence that are modelled after the information processing systems utilized by humans. Interconnections and processing elements are the two primary components that make up each ANN. Interconnections, also known as weights, are responsible for making connections between neurons, whereas the processing components, which can be neurons or nodes, are the ones responsible for processing information. MLP is still one of the most dominating structures of the ANN and also the structure with the most extended reach, despite the fact that the ANN has a wide variety of structures. The MLP is a universal function approximation that is utilized in the process of developing mathematical models through the utilization of regression analysis. This approximation is demonstrated by Cybenko's theorem (1989). The network is able to learn certain features concealed within the collected data samples through training on observation data, and it can even generalize what it has learnt from its training. MLP networks have a multilayered structure, with the first layer containing the data that will be used to train the model, the last layer containing the data that will be output by the model, and the layers in between the training data and the output data being referred to as hidden layers. The number of input variables is the same as the number of neurons in the input layer. The number of outputs is typically the same as the output parameter. The hidden layers are responsible for the internal appearance of the link between the model inputs and the output that is intended. The value of each neuron in the hidden layer, also known as the output layer, is calculated by adding the values of all of the neurons in the layer below it and multiplying that total by a weight that is specific to that neuron. After that, this value is added to the bias, and the resulting total is passed from an activation function. The structure of the MLP neural network that was used in this investigation may be seen in Figure 3, which shows that there were two hidden layers.

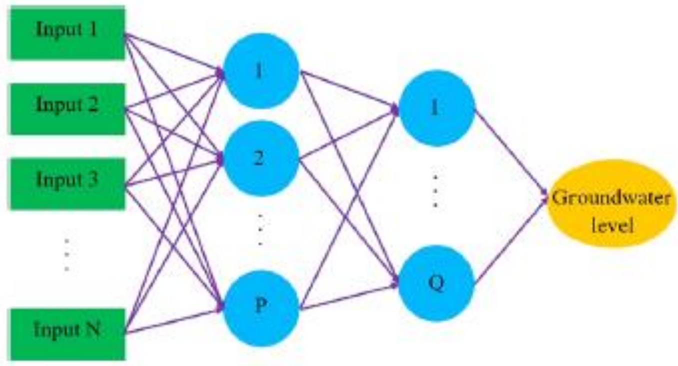

Figure 3: Schematic diagram of MLP neural network

The following is an overview of some of the activation functions that are typically used for hidden layers and output layers:

$$\text{Linear = Purelin: } f(x) = x \qquad (6)$$

$$\text{Logsig = Sigmoid: } f(x) = \frac{1}{1+e^{-x}} \qquad (7)$$

$$\text{Transig = Tanh: } f(x) = \frac{e^{x}- e^{-x}}{e^{x}+ e^{-x}} \qquad (8)$$

$$\text{Elliot 2 Symetric Sigoid = } f(x) = \frac{x^2 SIGN\ (x)}{1+x^2} \qquad (9)$$

The following is an example of the output of an MLP model with two hidden layers, the activation functions for which are Logsig and Tansig, while the Purlin activation function is used for the output layer.

$$\text{Output = Purelin } (w_3 * \left(Transig\left(w_2 * (Logsig(x) + b_1)\right) + b_2\right) + b_3) \qquad (10)$$

Where w1 and b1 are the weight matrixes and the bias vectors of the first hidden layers, w2 and b2 are the weight matrixes and the bias vectors of the second hidden layers, and w3 and b3 are the weight matrixes and the bias vectors of the output layers.

V. PROPOSED METHOD

PSO is a strong population-based evolutionary computation method. It takes its cues from the social behavior of fish and birds, such as fish schooling and bird flocking. PSO is used in a number of engineering and medical specialties [24]. Evolutionary approaches are used to solve the well-known travelling salesman problem (TSP), and the solutions were judged to be satisfactory [25]. On the basis of time-variant fuzzy time series, PSO is used for forecasting. PSO has found use in medicine for heartbeat classification and human tremor analysis.

The fuzzy approach for predicting groundwater level was optimized via PSO. But one difficult component of predicting groundwater level was reproducing the identical field conditions. The use of quantitative tools to anticipate outcomes is erroneous due to several practical uncertainties. As a result, data-driven models based on soft computing methods are now being used.

This section has examined the potency of data-driven models, such as FL and PSO. The goal of the study is to find a solution to the problem of choosing a length in fuzzy sets for fuzzy time series prediction. PSO is used to modify the lengths of fuzzy intervals in order to estimate groundwater levels.

Proposed Algorithm: For determining the length of fuzzy intervals, an FPSO approach is devised. The definition of a step-by-step process is as follows:

Step 1 - Define the Universe of discourse for inputs and outputs

As described in Chapter 3, the model for prediction is chosen by using groundwater recharge $R(t)$ and groundwater discharge $D(t)$ as inputs and groundwater level $W(t)$ as an output.

$$W(t) = f(R(t), D(t)) \qquad (11)$$

The $R(t)$, $D(t)$ and $W(t)$ have a range that changes between [0, 500], [1350, 2400], and [155, 170], correspondingly.

Step 2 - Initialization of Parameters

The Swarm size N, cognitive parameter $C1$, social parameter $C2$, inertia weight w, and maximum number of generations G_{max} are the PSO parameters that are employed.

From the interval of 10 to 50, $N = 30$ is chosen in steps of 10. $C_1 = 2$, $C_2 = 2$ (Haupt and Haupt [26], proposed to take the values of C_1 and C_2 as 2). From the range of 1.2 to 2.0, $w = 1.4$ is chosen in steps of 0.1. $G_{max} = 1000$, which is a stopping factor of appropriate size.

Step 3 - Construction of Fuzzy Sets

In mathematical modelling, triangular membership functions are frequently used because of their capability to construct a mediating solution. Divide the universe of discourse into 57 intervals that were built using PSO's random population generator, with X_1, X_2, …….. X_{19} describing R_t as illustrated in Figure 6, fuzzy sets are being used for the second input. Dt are illustrated in Figure 7 and the remaining fuzzy intervals, ranging from X_{39} to X_{57}, are transmitted to the output W_t, .

Step 4 - Construction of Rule base

The construction of the rule base model, which is done using the training dataset (Wang and Mendel). The Wang and Mendel method is employed, and it consists of five steps.

Step 5 - Forecast

The anticipated output for the fuzzy time series groundwater dataset is produced by MFIS through the use of the centroid approach.

Step 6 - Compare current objective value with pbid value

Perform a comparison of the present objective value of each particle with the value of its pbid. If the current value is superior, then the pbid for the particle should be updated to reflect the new objective value; otherwise, the pbid should be left unchanged.

Step 7 - Identify best particle of generation

Find the one particle out of the total population that has the best objective value and determine which one it is. If the objective value is higher than the gbid, then the gbid should be updated to reflect the objective value of the currently best particle [91].

Step 8 - Termination criteria

This procedure is carried out for the maximum possible number of generations, which is 1000. The highest number of generations is being used as the stopping criteria for this.

VI. Statistical Measures

The use of statistical methods is playing a significant part in the process of selecting the parameters that have been observed and computed as having the best overall fitness.

Root Mean Square Error (RMSE): The root mean square error (RMSE) is a typical statistical method that is used to calculate the difference between the values that are predicted by a model and the values that are actually gathered from the field (that is being modeled). In some circles, it is also referred to as the Root Mean Square Deviation (RMSD) [19]. Residuals are the individual differences that exist between every pair of predicted and observed values, and the RMSE combines all of these differences into a single measure of the accuracy of the prediction. The root mean squared error (RMSE) is defined as "square root of the mean squared error," which is described using the following relationship: The RMSE of a model prediction with respect to the estimated variable X_{model} is defined as "square root of the mean squared error."

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \tag{12}$$

Where X_{obs} is observed values and X_{model} is modeled values at time i. The RMSE values that are calculated have units, and these RMSE values can be utilized to differentiate between the performances of a model during a calibration period and a validation period. Additionally, it is helpful to compare the performance of the individual model to that of other predictive models.

Coefficient of Correlation (*R*): Linear correlation coefficient (*R*) measures the strength and the direction of a linear relationship between observed and predicted variables. The linear correlation coefficient is also referred as “Pearson product moment correlation coefficient” on the name of its developer Karl Pearson. The mathematical formula to compute linear correlation is stated by the following relationship

$$R = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \tag{13}$$

Where *n* represents the total number of individual data points. The range of values that are used to indicate the value of R is from $-1<R<+1$. The positive and negative signs represent positive linear correlations and negative linear correlations, respectively, between the values that were seen and those that were anticipated by the mathematical model or the soft computing models that were built [20].

Coefficient of Determination (R^2): It is the fraction of the overall variance that can be accounted for by the variations that have been explained. It provides the proportions of the variation of one variable in comparison to the variance that was anticipated for the same variable. It determines how definite a prediction can be based on the model results that have arrived. Its value, which can range from $0 < R^2 < 1$ and indicates the strength of the link between the observed variable and the predicted variable, can be anywhere in that range. It is a measure of the proportion of the data that lies along the line that provides the best fit [21]. R^2 is a statistic that determines how accurately the various regression lines depict the correlation that exists between the variables that were observed and those that were anticipated. In a linear regression plot, an R^2 value of 0.996 indicates that 99.6 percent of the arrived variance between the variables is explained by the linear relationship, while the remaining percentage is unexplained. If R^2 = 0.998, however, only 90.8 percent of the variance is explained by the linear relationship.

The following mathematical relationships give rise to the mathematical formulation of R^2:

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} \tag{14}$$

Mean Square Error (MSE): The mean squared error, often known as MSE, is the average of the squares of the error, with error being defined as the difference between the anticipated and observed values in the dataset. The formula derived from mathematics a computation of such mean square error is carried out utilizing the relationship that follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{15}$$

Mean Absolute Error (MAE): In time series analysis, this will help to measure the predicting error, which permits error identification in the final outcomes. The mean absolute error (*MAE*) is calculated by adding up all of the absolute values of the residual and counting the average difference [27]. The mathematical representation of such a measure can be derived by computing it using the following relationship:

$$MAE(t) = SAE/N = \sum_{i=1}^{N} \frac{x_{pre} - x_{obs}}{N} \tag{16}$$

Where x_{pre} represents the parameters that were predicted, x_{obs} represents the parameters that were actually observed, *SAE* stands for the sum of absolute error, and *N* represents the number of data samples.

VII. Simulation Results

In the Tiruvallur district of Tamil Nadu State, fuzzy, ANN, and PSO models have been built to estimate the water level by applying the database of 192 months of test data from 2004 to 2020 taken from the Kotralai River. This data will be used until 2020. An artificial neural network (ANN) with a 1-2-3-layer back propagation structure, five input nodes in the input layer, and two hidden layers has been selected for the development of the models.

The data from four reservoirs 'POONDI', 'CHOLAVARAM', 'REDHILLS', 'CHEMBARAMBAKKAM' located in Chennai Tamil Nadu has been taken for the simulation of this project. Figure 4 shows all data and shows all rainfall data.

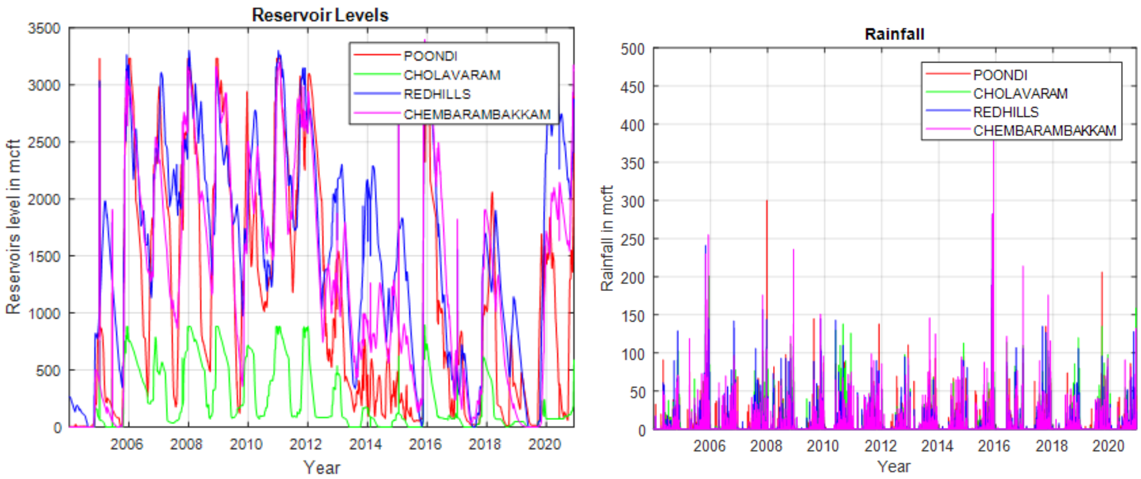

Figure 4: Four reservoirs level and rainfall data

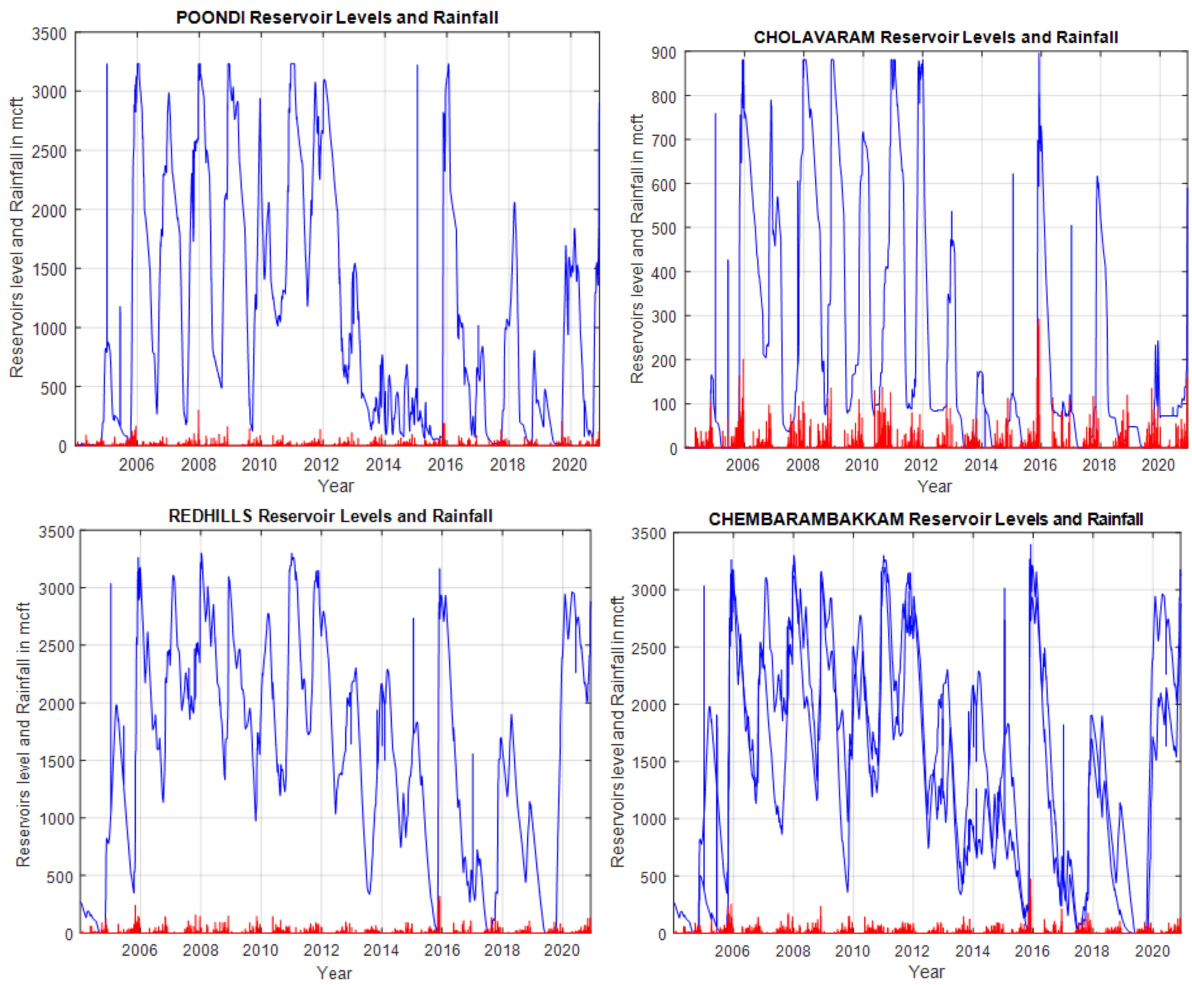

Figure 5: Four reservoir level and rainfall data

FIS is the procedure of mapping specified inputs to set of outputs using FL. FIS are of two types – Mamdani-FIS and Sugeno-FIS. All basic FL models are developed using Mamdani-FIS. Working of Mamdani-FIS includes fuzzification of inputs, construction of rule- base and defuzzification to get crisp output.

Due to the fact that the data for this study are in qualitative form and not fuzzy form, fuzzification was utilised so that inputs and outputs could be converted into the fuzzy form. Figures 6 through 8 show that W_t is alienated in seven fuzzy subsets (EL, VL, L, M, H, VH, EH), that D_t is alienated in six fuzzy subsets (VL, L, M, H, VH, EH), and that R_t is alienated in five fuzzy subsets (VL, L, M, H). In this particular investigation, a triangle membership function (MF) was utilised. The membership grades are determined with the assistance of the data and under the guidance of the experts.

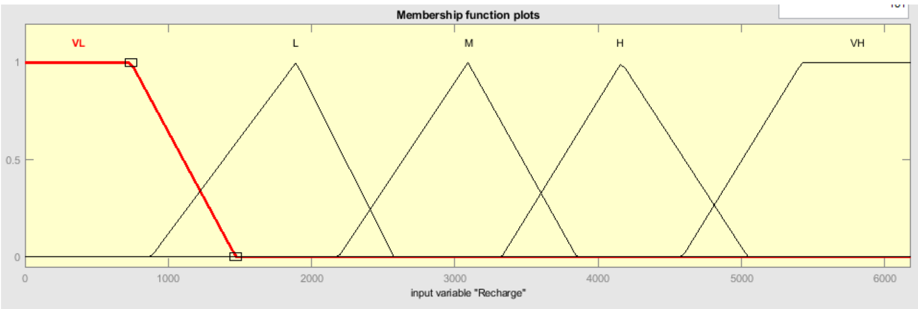

Figure 6: Fuzzy subsets of R_t

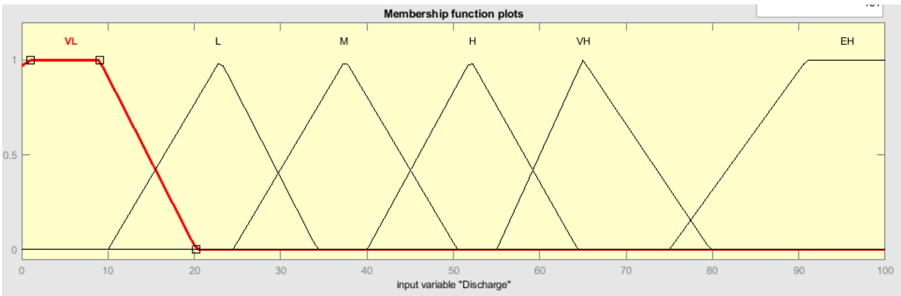

Figure 7: Fuzzy subsets of D_t

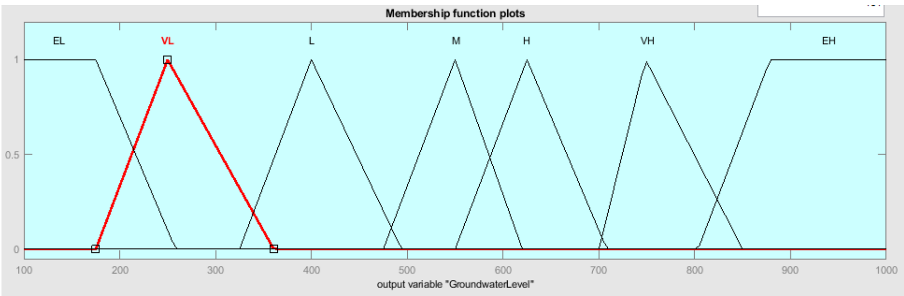

Figure 8: Fuzzy subsets of W_t

Rule Base

A collection of IF-THEN rules is what the fuzzy rulebase is made out of. Rules are derived from the numerical data that has been observed over the course of the previous years and describes the level of groundwater in terms of both groundwater recharge and groundwater discharge. Table 1 shows all the rules taken in to develop the model.

Table 1: Fuzzy Rule Base Table

Recharge R_t	**Discharge D_t**	**Water Level W_t**	**Recharge R_t**	**Discharge D_t**	**Water Level W_t**
VL	VL	VL	H	M	M
VL	L	L	H	VH	VL
VL	H	L	H	EH	EL
VL	M	L	M	VL	L
VL	VH	VL	M	L	M
VL	EH	VL	M	H	L
L	VL	L	M	M	M
L	L	L	M	VH	VL
L	H	EL	M	EH	EL
L	M	VL	VH	VL	H
L	VH	VL	VH	H	M
L	EH	EL	VH	M	H
H	VL	M	VH	VH	VH
H	L	EL	VH	EH	EH
H	H	H			

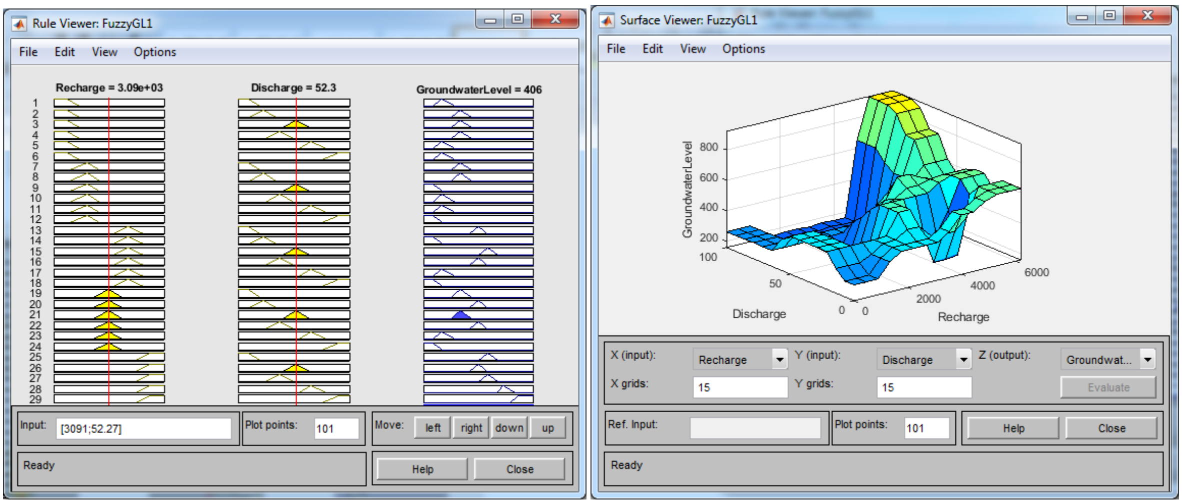

Figure 9: Fuzzy Rule Viewer and Surface Viewer

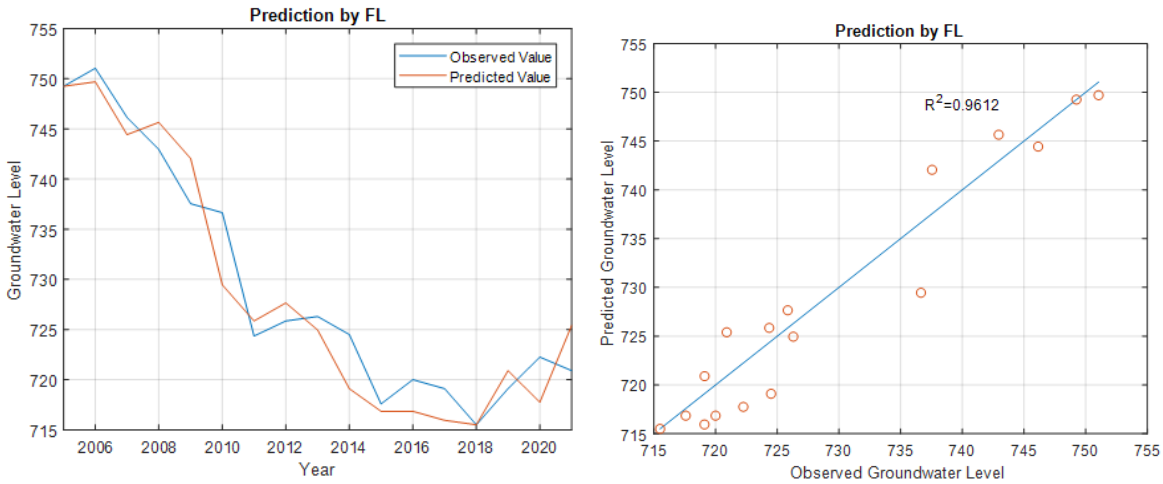

Figure 10: Observed and predicted values using FL

The Coefficient of Determination (R^2) is 0.9612.

Artificial Neural Networks (ANN) models: The models are developed for one, two and three hidden layers. In case of two layers (hidden) models, quantity of neurons are selected using trial and error. These ANN models are trained and tested using **nntool** in MATLAB. In this study, back-propogation with Bayesian regulation backpropagation technique was used to train the ANN model. The sigmoid transfer function, 'tansig', log sigmoid 'logsig' and pure linear 'purelin' has been choosen as transfer functions for both the hidden layer and output.

The ANN1, ANN2, ANN3 (with two hidden layer) models performed better than respectively one hidden layer models while for ANN 3 and ANN 5 (one hidden layer) models performed better than respective two hidden layer models. Further, best five ANN models are compared to respective FL models as shown in Figures 11.

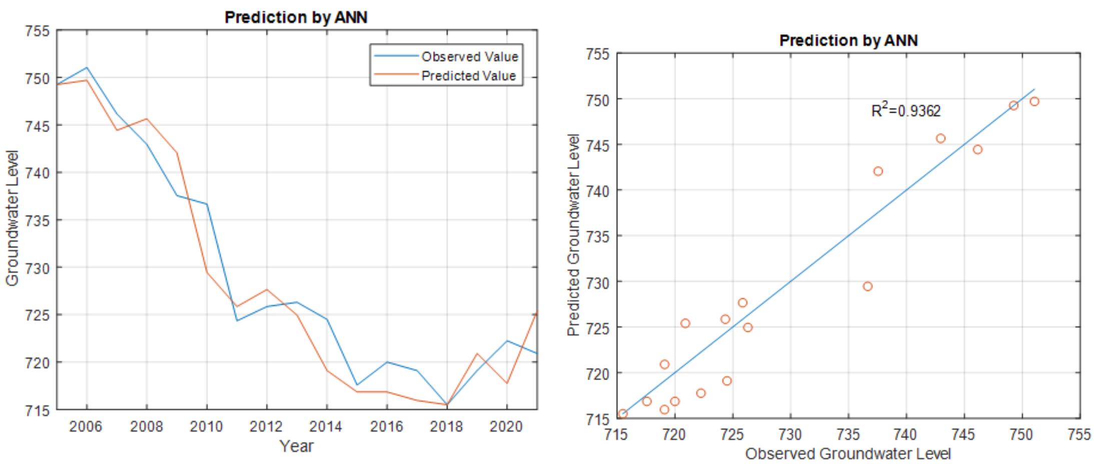

Figure 11: Observed and predicted values using ANN

The Coefficient of Determination (R^2) is 0.9362.

Fuzzy PSO (FPSO) model: The PSO algorithm is run for a total of five times for each possible circumstance before the best result is saved. At the beginning stage, various values of the algorithm parameters are tried, and the results were found to be most effective for *N* = 30, the size of the swarm, *C1* = 2, the cognitive parameter, *C2* = 2, the social parameter, and *w* = 1.4.

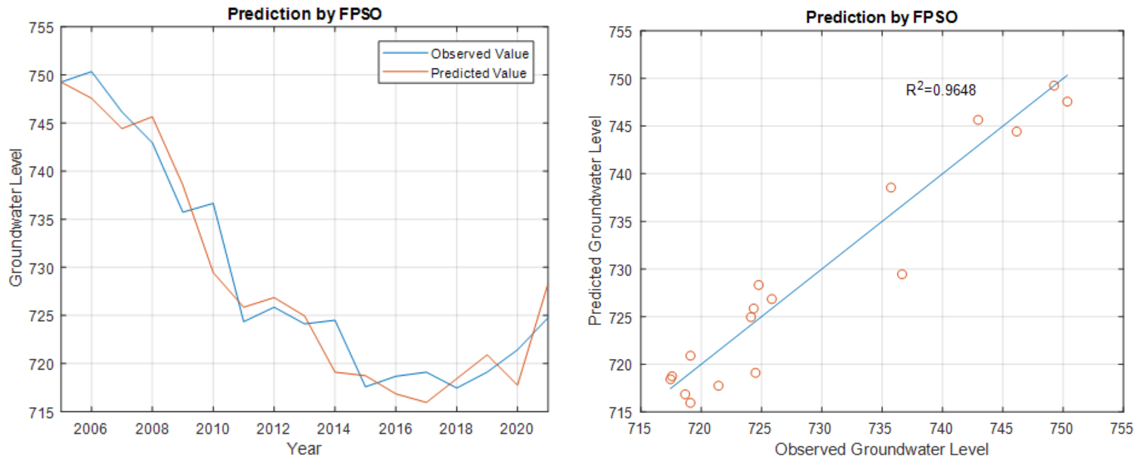

Figure 12: Observed and predicted values using FPSO

The Coefficient of Determination (R^2) is 0.9648.

In our earlier research for the purpose of forecasting groundwater level, we took into account historical data from the Chennai District pertaining to groundwater recharge, groundwater outflow, and groundwater level during the period of time between 2004 and 2020. In this investigation, the FPSO approach makes use of the same data. In Table 2, which depicts comparisons of results obtained through the use of various methods, an illustration is provided of a comparison of the FPSO method with the FL and ANN methods respectively. Every approach is evaluated using the same standard set of parameters. MSE and Coefficient of Determination are the criteria that are utilized in the process of performing an efficiency assessment on the FPSO method.

Table 2: Performance indicators for FL, ANN & FPSO models

Model	Hidden Layer	MAD	R	R^2	RMSE
FL1	N.A.	0.350	0.975	0.9612	0.428
ANN1	One	0.575	0.947	0.9050	0.719
ANN2	Two	0.455	0.943	0.9360	0.612
FPSO	N.A.	0.411	0.945	0.9648	0.2737

VIII. CONCLUSION

In the beginning, FL and ANN were utilized in order to predict the amount of groundwater. The experiment evaluated FL and ANN for predicting groundwater level using various combinations of inputs for recharge and outflow. In order to examine, these artificial neural network (ANN) models were compared (three for each one and two hidden layers), and the best of the two ANN models were compared to the corresponding five FL models. The R^2 values for ANN models ranged from 0.9050 to 0.6360, and the R^2 values for FL models started at 0.9612 and for FPSO model is 0.9648. Based on the evaluation criteria, the results that ANN2 (two hidden layers) produced were superior to those produced by any other developed ANN models. The FL fared better than any of the ANN models. In addition, the present research recognized that FPSO produced better outcomes for a reduced number of inputs, and that ANN gave results that were encouraging.

REFERENCES

[1] Afzaal, H., Farooque, A. A., Abbas, F., Acharya, B., & Esau, T. (2020). Groundwater estimation from major physical hydrology components using artificial neural networks and deep learning. Water, 12(1), 5.

[2] Asefa T, Kemblowski MW, Urroz G, McKee M, Khalil A (2004), Support vectors-based groundwater head observation networks design. Water Resour Res 40:1–14

[3] Balavalikar S, Nayak P, Shenoy N, Nayak K (2018) Particle swarm optimization based artificial neural network model for forecasting groundwater level in Udupi district. AIP Conf Proc 1952(1):020021

[4] Fallah-Mehdipour E, Bozorg Haddad O, Marin˜o MA (2013) Prediction and simulation of monthly groundwater levels by genetic programming. J Hydro Environ Res 7:253–260

[5] Feng S, Kang S, Huo Z, Chen S, Mao X (2008) Neural networks to simulate regional ground water levels affected by human activities. Ground Water 46:80–90

[6] Kouziokas GN, Chatzigeorgiou A, Perakis K (2018) Multilayer feed forward models in groundwater level forecasting using meteorological data in public management. Water Resour Manag 32(15):5041–5052

[7] L. A. Zadeh, "Foreword", in Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks", Iizuka, Japan, pp. XIII-XIV, 1992.

[8] W. S. McCulloch and W. Pitts,"A logical calculus of the ideas immanent in nervous activity", Bull. Math. Biophys., vol. 5, no. 4, pp. 115–133, 1943.

[9] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology," Artificial neural networks in hydrology. II: Hydrologic applications", Journal of Hydrologic Engineering, vol. 5, no. 2, pp.124-137, 2000.

[10] B. Bhattacharya and D. P. Solomatine, "Application of artificial neural network in stage-discharge relationship", in Proceedings of the 4th Int. Conf. on Hydro- informatics. Iowa, USA, 2000.

[11] D. C. S. Bisht, M. M. Raju and M. C. Joshi, "ANN based river stage-discharge modelling for Godavari river, India", Computer Modelling and New Technologies, vol 14, no. 3, pp. 48-62, 2010.

[12] M. Mitchell, An Introduction to genetic algorithms, Cambridge, MIT Press, 1996.

[13] D. E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, Delhi, Pearson Education, 2005.

[14] D. E. Goldberg and K. Deb, “A comparison of selection schemes used in genetic algorithms”, in Proceedings of Foundations of Genetic Algorithms - I, San Mateo, Ed. by G. Rawlins, Morgan Kaufmann, pp. 230-236, 1991..

[15] A. Zadeh,“From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions”, International Journal of Applied Math and Computer Science, vol. 12, no. 3, pp. 307-324, 2002.

[16] A. Bardossy, I. Bogardi and L. Duckstein, “Fuzzy regression in hydrology”, Water Resour. Res., vol. 25, no. 7, pp. 1497–1508, 1990.

[17] A. Bardossy, R. Hagaman, L. Duckstein and I. Bogardi,“Fuzzy least squares regression: theory and application”, in Proceedings of Fuzzy Regression Models, Eds. M. Fedrizzi, and J. Kacprzyk, pp. 66–86, 1991.

[18] E. C. Ozelkan and L. Duckstein,“Multi-objective fuzzy regression: a general framework”, Computers and Operations Research, vol. 27, no. 7–8, pp. 635–652, 2000.

[19] Y. Hundecha, A. Bardossy and H.W. Theisen,“Development of a fuzzy logic- based rainfall-runoff model”, Hydrol. Sci. J., vol. 46, no. 3, pp. 363–376, 2001.

[20] S. P. Simonovic,“ Reservoir systems analysis: closing gap between theory and practice”, J. Water Resour. Plan Manage ASCE, vol. 118, pp. 262–280, 1992.

[21] S. Jain, P. C. Mathpal, D. Bisht and P. Singh,“A unique computational method for constructing intervals in fuzzy time series forecasting”, Cybernetics and Information Technologies, vol. 18, no. 1, 2018.

[22] P. Bortolet. “Modelisation et commande multivariables floues: application a la commande d’un moteur thermique”. Ph.D. Thesis (unpublished), Toulouse, INSA, 1998.

[23] P. H. Sneath,“ The application of computers to taxonomy”, J. Gen. Microbiology, vol. 17, pp. 201-226, 1957.

[24] S. Alvisi, G. Mascellani, M. Franchini and A. Bardossy,“ Water level forecasting through fuzzy logic and artificial neural network approaches”, Hydrology and Earth System Sciences Discussions, vol.10, no. 1, pp. 1-17, 2006.

[25] A. K. Affandi and K. Watanabe,“Daily groundwater level fluctuation forecast- ing using soft computing technique”, Nature and Science, vol. 5, no. 2, pp.1-10, 2007.

[26] , K. Solaimani,“Rainfall-Runoff prediction based on artificial neural network (A case study: Jarahi Watershed)”, American-Eurasian J. Agric. and Environ. Sci, vol. 5, no. 6, pp. 856-865, 2009.

[27] N. Fernandez, W. Jaimes and E. Altamiranda,“Neuro-fuzzy modeling for level prediction for the navigation sector on the Magdalena River (Colombia)”, Journal of Hydroinformatics, vol. 12, no. 1, pp. 36-50, 2010.

[28] M. K. Mayilvaganan, K. B. Naidu,“ ANN and fuzzy logic models for the pre- diction of groundwater level of a watershed”, International Journal on Computer Science and Engineering, vol. 3, no. 6, pp.2523-2530, 2011