ISSN:2229-6107

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY

E-mail :
editor.ijpast@gmail.com
editor@ijpast.in

www.ijpast.in

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

GROUND WATER LEVEL PREDECTION USING HYBRID ARTIFICUAL NEURAL NETWORK WITH GENETIC ALGORITHM

1.K VENKATESH,2. B. SAI RUCHITHA,3. B. VINEELA,4. S. SUSHMA,5. S. ANJALI

ABSTRACT:
In recent years, the growth of the economy has led to the increasing exploitation of water resources and groundwater. Due to heavy abstraction of groundwater its importance increases, with the requirements at present as well as in future. Accurate estimates of groundwater level have a valuable effect in improving decision support systems of groundwater resources exploitation. This paper investigates the ability of a hybrid model of artificial neural network (ANN) and genetic algorithm (GA) in predicting groundwater levels in an observation well from Udupi district. The ground water level for a period of ten years and rainfall data for the same period is used to train the model. A standard feed forward network is utilized for performing the prediction task. A groundwater level forecasting model is developed using artificial neural network. The Genetic Algorithm is used to determine the optimized weights for ANN. This study indicates that the ANN-GA model can be used successfully to predict groundwater levels of observation well. In addition, a comparative study indicates that the ANN-GA hybrid model performs better than the traditional ANN back-propagation approach.

1.INTRODUCTION

1.1. RESEARCH BACKGROUND

Groundwater resources, as one of the most valuable and important sources of water in the world, play a direct and crucial role in various aspects of human lives, such as agriculture, industrial development, and potable water supply [1,2]. In addition, the indirect effects of groundwater resources on the environment and communities are undeniable. The groundwater level (GWL) is a direct and simple measure of groundwater availability and accessibility. Having a proper understanding of the past, current, and future situations of GWL can provide policy-makers and practitioners in water sectors with better insight and perception to develop strategies for the planning and management of water resources, to ensure sustainable socioeconomic development [2]. However, GWL consists of an integrated response to several climatic, topographic, and hydrogeological factors and their interactions, which makes the simulation of GWL a challenging task [3,4]. Numerous studies using different simulation approaches have been conducted for the quantitative and qualitative prediction of GWL. These methods cover a wide range of physically based

1.ASSISTANT PROFESSOR,2,3,4&5 UG SCHOLAR
DEPARTMENT OF IOT, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

INTERNATIONAL JOURNAL OF
PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST
ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

conceptual models, experimental models [5–7], and numerical models. Modeling groundwater using numerical models consists of several approaches, such as finite difference [8], finite volume [9], finite element [10], and element-free [11] methods. Even though these classical models are robust and reliable, the precision and accuracy of numerical models are confined by several factors, such as their high dependency on large volumes of data related to aquifer properties, the geology of the porous media, and basement topography [12]. Moreover, properly demarcating domain boundaries, defining an efficient grid size for solving the associated differential equations, and calibrating/validating the executed model have made numerical modeling a complex and sophisticated task. In last two decades, artificial intelligence (AI) models have been widely used to overcome the drawbacks of conventional numerical models for GWL simulation. Fig. 1 presents the goal map, depicting the two major pieces of information, one being the most studied geographical locations and other those which have not yet been

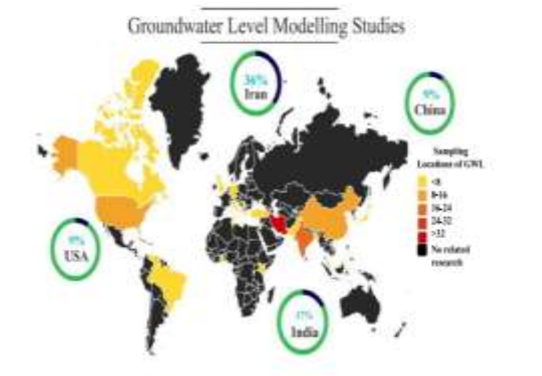


Fig. 1. Map representation of sampling location of GWL data all over the globe with specified area with no related research on GWL modeling using AI models.

studied. Furthermore, Fig. 1 highlights the four major countries which have done extensive GWL modeling-related studies, whereas the black color zone reveals the areas where the application of AI has not yet gained in popularity. Around 70% of areas have not yet used GWL, as many do not need GW-related studies, due to a sufficient amount of surface water or less habitants, such as in polar areas, Russia, and so on. Moreover, some underdeveloped countries, such as Africa, and some parts of Asia and North America, may not have explored AI techniques yet. As per Fig. 2, there has been a significant increase in studies in this field in the last few years; however, more studies should be done, based on different geographical locations, to test the efficiency of the proposed models. The usability and reliability of AI models in dealing with complex and high-dimensional engineering problems have been proven in the last few decades [13–15]. AI consists of multidimensional systems combining various mathematical and statistical components and arithmetic and heuristic algorithms. AI has been extensively employed in different fields of science, engineering design, energy, robotics, and economics . It has also been intensively used for solving various civil and environmental engineering problems. Some examples include soft computing techniques, Machine Learning (ML) methods, probabilistic analysis, and Fuzzy-based systems. In recent years, more attention has been paid to the successful use of AI in different hydrological fields, including water resources, surface and groundwater hydrology, sediment contamination, and hydraulics.

1.2. Research significance Proper measurement, nowcasting, and forecasting of GWL in aquifers are highly important for the sensible management of groundwater resources. Monitoring GWL can provide hydrologists and hydrogeologists valuable information to understand the short- and long-term variations in groundwater availability. The ability of AI models to simulate and predict GWL without requiring deep and comprehensive knowledge of the underlying topographical and hydro-

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

geophysical parameters makes them appealing methods compared to physically based and numerical methods

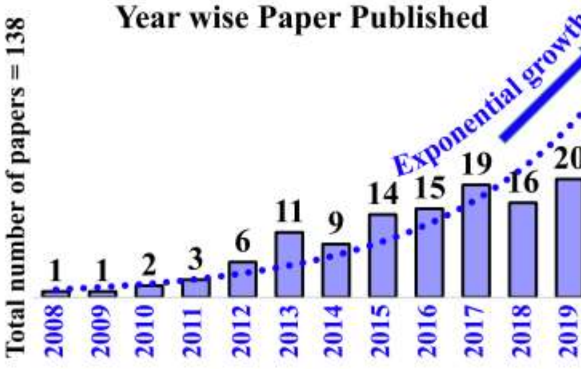


Fig. 2. Arithmetical conceptualization of growth observed in GWL researc based model during 2008–2020.

A large volume of studies have already investigated and reported the applicability of AI in modeling GWL over the last two decades. Most of the early works included simple and standard AI methods, such as perceptron Artificial Neural Networks (ANNs). However, in the last decade, the application of a variety of ML models for GWL simulation has been witnessed; examples include different types of ANNs, fuzzy-based models, Support Vector Machines (SVMs), tree-based models, Genetic Programming (GP), and Gene Expression Programming (GEP) models. Most recently, along with the application of novel AI models, including Deep Learning (DL), Extreme Learning Machine (ELM), and Long Short-Term Memory (LSTM) , novel strategies, such as integrated and hybrid AI models , ensemble learning, and AI-GIS (Artificial Intelligence-Geographic Information System)-based models, have been implemented for modeling GWL. Rajaee et al., for instance, studied 67 journal papers and provided a bibliographic review of the applications of AI in GWL simulation and forecasting. Considering the outcomes of different classic AI methods, such as ANNs, Adaptive Neurofuzzy Inference System (ANFIS), SVM, GP, and hybrid AI methods, the study concluded that AI methods can be successfully used to model and forecast GWL in aquifers located in regions with different geology and climate. Some studies have attempted to combine the advantages of AI and numerical methods to develop hybrid models. For example, Nourani and Mousavi introduced a hybrid AI-meshless model for modeling GWL. They used AI methods, such as ANN and ANFIS, for temporal modeling of GWL, while the meshless method was used for solving the governing differential equations to estimate the GWL in places with no observations . Chen et al. carried out a comparative study using a finite difference numerical model versus three ML models, including ANNs and SVM, for simulating GWL. Comparing the general performance of the two distinct approaches revealed that the ML models acted better than the numerical model. Nevertheless, they also mentioned the superiority of the finite difference method, due to its generalization ability in including the physical mechanism of the aquifer.

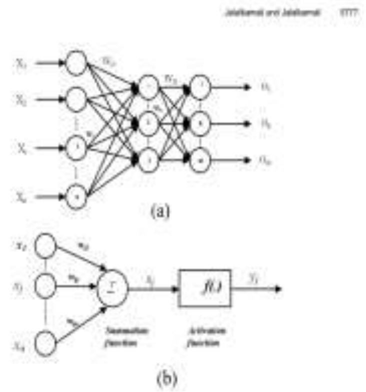
ALGORITHMS Artificial neural networks Artificial neural networks estimation approach has received tremendous attentions in the last few decades. An interesting property of ANNs is that they often work well even when the training data sets contain noises and measurement errors (Hammerstrom, 1993). Moreover, they have the capability of representing complex behaviors of nonlinear systems (Maier and Dandy, 2000). The advantage of the ANN is that with no prior knowledge of the actual physical process and, hence, the exact relationship between sets of input and output data, if acknowledged to exist, the network can be trained to learn such a relationship. The ability to train and learn the output from a given input makes ANN capable of describing large scale arbitrarily complex non-linear problems. A neural network is characterized by its architecture that represents the pattern of connection between nodes, its method of determining the connection weights, and the activation

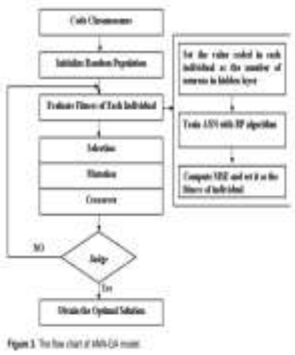
ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

function (Fausett, 1994). A typical ANN consists of a number of nodes that are organized according to a particular arrangement. Feed forward neural network models One way of characterizing ANNs is based on the direction of information flowing and processing, as feed-forward (where the information flows through the nodes from the input to the output side) and recurrent (where the information flows through the nodes in both directions). Among these combinations, the multi-layer feedforward networks, also known as multi-layer perceptron (MLPs), trained with a back-propagation learning algorithm have been found to provide the best performance with regard to input-output function approximation, such as forecasting applications. A typical MLP with one hidden layer is shown in Figure 1; (a). The first Layer connects with the input variables and is called the input layer. The last layer connects to the output variables and is called the output layer. The layer between the input and output layers, is called the hidden layer (there may be more than one hidden layer in an MLP).The processing elements in each layer are called nodes or units. Each node is connected to the nodes of neighboring layers. The parameters associated with each of these connections are called weights. The architecture of a typical node (in the hidden or output layer) is also shown in Figure 1; (b) Each node j receives incoming signals from every node i in the previous layer. Associated with each incoming signal xi is a weight wji. The effective incoming signal sj to node j is the weighted sum of all the incoming signals, is passed through the effective incoming signal, sj non-linear activation function (sometimes called a Transfer function or threshold function) to produce the outgoing signal yj of the node. j j s f s 1 exp 1 ( ) The most commonly used function in an MLP trained with backpropagation algorithm is the sigmoid function. The sigmoid function most often used for ANNs is the logistic function (Sivakumar et al., 2002): n i S j wji xi 0 (2) Recurrent neural network models The recurrent neural network (RNN) is another multi-layer architecture that has been used for a variety of applications including control systems and forecasting of dynamic processes. In this section RNN structure is briefly discussed. The RNN architecture, a variation of general feed-forward backpropagation (FFBP) architecture, is used to capture dynamic and highly nonlinear systems by including a feedback mechanism in the architecture. The general RNN architecture uses specialized hidden nodes to introduce feedback to the network. In such a network, the output of these specialized nodes is provided as input to others. Once such feedback connections are allowed, the network topology becomes more connected since any node can be connected to any other node, including to itself. These self-connected or selfrecurrent feedback nodes form the "context" layer of a network and are tagged on to the network structure along with the usual nonfeedback nodes. The "context" layer is used to retain information between training iterations and serve as memory of the system by retaining the state of the network before the next set of data is processed. Each time a pattern is presented, each context node computes its activation just as in a feed forward network. However, its output is now able to reflect the state of the network before the pattern is seen. When subsequent patterns are presented, the hidden and output units' states will be a function of everything the network has seen so far. Thus, at each time period, activation propagates only forward through one layer of connections. Once some level of activation is present in the network, it will continue to flow through all the remaining hidden layers, even in the absence of any new input whatsoever. However, this added feedback mechanism (memory function) requires additional network connections, a large amount of

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST
ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

storage and computation, and a larger training set in order for the RNN to work well. This invariably leads to difficult network training and slow convergence (Atya and Parlos, 2000). Training methodology in RNN The method of RNN training is similar to that of feed forward network models. The training algorithm is explained with the help of a simple example. A small network which has two input neurons, one hidden layer having three neurons and one output neuron is shown in Figure 2. In addition, a neuron taking input from the output layer and connected to the hidden layer is added as shown. This neuron is the additional neuron in RNN.





Genetic algorithm GA optimizes using a search process that emulates natural evolution. On the other hand GA is a global heuristic, stochastic, optimization technique based on evolution theory and genetic principles developed by (Holland, 1975). Goldberg and Michalewicz (1992) discussed the mechanism and robustness of GA in solving nonlinear optimization problems (Goldberg, 1989; Michalewicz, 1992). The algorithm begins with a randomly generated population which is consisting of chromosomes, and applies three kinds of genetic operators: The selection, crossover and mutation operators to find the optimal solutions. The selection operator chooses chromosomes from the current population based on fitness value of the individuals. The crossover operator combines the features of two parent chromosomes to form two similar offspring by swapping corresponding segments of the parents (Goldberg, 1989). The mutation operator creates new chromosomes by randomly changing the genes of existing chromosomes. GA can explore the entire design space by the genetic manipulations; it does not easily fall into a certain local minima or maxima. As this occurs, the GA converges to increasingly better solutions. Improvements in fitness, however, diminish as the population diversity decreases and the population converges toward a good solution. Stopping criteria such as "100 generations without improvement" and minimum population diversity are often used to terminate the algorithm when improvements are sufficiently small and infrequent. These concepts are well described in (Davis, 1991; Goldberg, 1989). Therefore, GA is an aggressive search technique that quickly converges to find the optimal solution in a large solution domain. ANN-GA model scheme In this research, a multi-layered feed-forward neural network (FFN) and recurrent neural network (RNN) with a back propagation algorithm are adopted. Although the back propagation algorithm is successful, it has some disadvantages. The algorithm is not guaranteed to find global minimum of error space and the convergence tends to be extremely slow. In addition, the selection of the learning factor and inertial factor affects the convergence of the BP

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

neural network which is usually determined by experience. In present research, the number of neurons in the hidden layer is determined using the genetic algorithm. The number of hidden layers and the number of nodes in each layer depends on the complexity of the patterns and the nature of the problem to be solved. The use of a single hidden layer is sufficient to approximate to any continuous function as closely as requested (Funahashi, 1989; Hornik et al., 1990) and studies also showed that having more than two layers may not result in significant performance improvements (Patuwo et al., 1993) Thus, in our study, a two-layer ANN is utilized (Figure 1). The number of neurons in the input and output layers are given by the number of input and output variables of network. The number of neurons in hidden layer is obtained by GA. In this study, an ANN with one hidden layer is employed. The number of neurons in this layer is determined by GA. The optimization process flow chart of the ANN-GA model is shown in Figure 3. The sigmoid function was used in each node of the hidden layer and output layer as the transfer function. Number of neurons in hidden layer is the only information that is coded in a chromosome in GA. After that, the GA is run and in its fitness assignment past, an ANN which the number of its hidden layer neuron is determined by coded chromosome is trained via ANN. Then the MSE of this trained ANN is set as the fitness values. The GA will generate many of individual values which they will be set to MSE. This process is depicted in Figure 3. Simulation setup Study area and data set The area which studied in this research is the Kerman plain aquifer which is a part of Kerman province located in the south-eastern of Iran as shown in Figure 4. In this plain, no permanent river exists; therefore, the supply of water demands in agriculture, industry, Jalalkamali and Jalalkamali 5779 Figure 4. The location of wells in Kerman plain. Figure 5. Time series plot for the rainfall versus month. domestic and municipal sectors in 3200 km2 area around this plain highly depends on groundwater. In the past two decades, frequent hydrologic droughts besides the increasing number of pumping wells have caused a decline rate of 1 to 3 m annually. As a consequent the groundwater quality has decreased as well. The long-term annual precipitation for the area has noticeably decreased from 150 to 100 (mm/year) during the 20 past years (1988 to 2009). The data acquired from the area consists of rainfall depth, temperature and depth of the wells time series measured at Kerman airport station (latitude 30o , 16' N, longitude: 56o , 54’ E). The data set was provided by Iranian Ministry of Energy (IMOE). The time series used in this research are summarized for a 22 year period (1988 to 2009). Figure 5 presents, the monthly precipitation at meteorological Kerman airport station. In this region most of annual rainfall occurred during the winter season. Because 5780 Afr. J. Agric. Res. Figure.6. Time series plot for the temperature versus month Figure 6. Time series plot for the temperature versus month. Table 1. The monthly statistical parameters of data. Data set Unit Xmean Sx Csx Xmax Xmin HNO.26 m -33.36 1.74 0.63 -28.82 -35.93 HNO.16 m -37.71 4.3 -0.39 -30.35 -45.49 HNO.41 m -34.57 3.72 0.43 -26.11 -40.53 R mm 11.26 17.39 2.3 109.1 0 T c o 15.82 7.68 0.03 29.25 1.05 of relatively high temperature of this province, temperature plays an important role in the water budget. Figure 6 shows the monthly temperature for the period mentioned. The data sample consisted of 22 years (1988 to 2009) of monthly records of air temperature (T), rainfall (R) and water levels in target well (H NO.26) and neighboring wells (H NO.16 and H NO.41). The first 19 years (1988 to 2006) data were used to train the models and the remaining data for testing. The monthly statistics of each time series are given in Table 1. In the table the Xmean, Sx, Csx, Xmax and Xmin respectively denote the mean, standard deviation, skewness

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

coefficient, maximum and minimum of observations. Parameter setup Population size and generation numbers are set to 100. The tournament selection is used as selection method in GA, two point crossover and an uniform mutation are consider for reproduction Crossover rate and mutation probability are set to 0.7 and 0.01 respectively. Learning rate in BP algorithm is set to 0.02 and 50 epochs are considered for training the ANN.

CONCLUSIONS:

With climate change and overexploitation situations, groundwater table fluctuations' accurate predictions are essential for managing groundwater resources. The present study aimed to investigate the comparative potential of the hybrid GA-ANN models against the traditional GA models to predict the seasonal groundwater table depth in the area between the Ganga and the Hindon rivers. The ability of developed models was evaluated by using the statistical indicators (coefficient of determination, coefficient of efficiency, correlation coefficient, mean absolute deviation, root mean square error, coefficient of variation of error residuals, absolute prediction error, and performance index), as well as through visual inspection. The analysis results demonstrate that the GA models recognized the groundwater table depth trend efficiently but failed to predict the groundwater table depth because the maximum coefficient of determination was only 0.47. Simultaneously, the GA-ANN models' performance was found to be superior to the GA models for GWTD prediction in both the seasons, with the highest coefficient of determination values of 0.94 and 0.95, respectively. It was also concluded that the more significant number of input parameters enhanced the predictive rationality of applied GA-ANN models. Thus, the GA-ANN based models may be successfully functional in the field of groundwater to predict the groundwater table fluctuations with reasonably good accuracy. The efficient models found in this study confirm promising outcomes and proved to be reliable and time-saving technologies for optimal planning and management of groundwater resources in the study area. Our proposed model could be readily transferable or adapted to other areas, specifically those with similar hydrogeological conditions. The accessibility and quantity of data are challenging. In future research, the authors will project to establish a wireless sensor network for near real-time monitoring of groundwater levels and meteorological data in the study area. Author Contributions: Conceptualization, K.P. and A.M.; methodology, K.P.; software, K.P.; validation, K.P., A.M., S.K. and A.K.; formal analysis, K.P., and A.M.; investigation, K.P., A.M., S.K. and A.K.; data curation, K.P.; writing—original draft preparation, K.P., A.M., and A.K.; writing—review and editing, K.P., A.M., S.K. and A.K.; visualization, K.P., A.M., S.K. and A.K.; supervision, A.M., S.K. and A.K.; project administration, A.K.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript. Funding: This research received no external funding. Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES:

1. Nunno, F.D.; Granata, F. Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network. Environ. Res. 2020, 190, 110062. [PubMed].

2. Amarasinghe, U.A.; Smakhtin, V. Global water demand projections: Past, present and future. Int. Water Manag. Inst. (IWMI) Colombo Sri Lanka 2014, 156, 1–24.

3. Haas, J.C.; Birk, S. Characterizing the spatiotemporal variability of groundwater levels of alluvial aquifers in different

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse 4. Nov 2023

settings using drought indices. Hydrol. Earth Syst. Sci. 2017, 21, 2421–2448. [CrossRef].

4. Yu, H.; Wen, X.; Feng, Q.; Deo, R.C.; Si, J.; Wu, M. Comparative Study of Hybrid-Wavelet Artificial Intelligence Models for Monthly Groundwater Depth Forecasting in Extreme Arid Regions, Northwest China. Water Resour. Manag. 2018, 32, 301–323. [CrossRef].

5. Goldman, M.; Neubauer, F.M. Groundwater exploration using integrated geophysical techniques. Surv. Geophys. 1994, 15, 331–361. [CrossRef].

6. Singh, A.; Malik, A.; Kumar, A.; Kisi, O. Rainfall-runoff modeling in hilly watershed using heuristic approaches with gamma test. Arab. J. Geosci. 2018, 11, 1–12. [CrossRef]

7. Malik, A.; Tikhamarine, Y.; Souag-Gamane, D.; Kisi, O.; Pham, Q.B. Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. Stoch. Environ. Res. Risk Assess. 2020. [CrossRef].

8. Tikhamarine, Y.; Souag-Gamane, D.; Ahmed, A.N.; Sammen, S.S.; Kisi, O.; Huang, Y.F.; El-Shafie, A. Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. J. Hydrol. 2020, 589, 125133. [CrossRef].

9. Malik, A.; Kumar, A.; Singh, R.P. Application of Heuristic Approaches for Prediction of Hydrological Drought Using Multi-scalar Streamflow Drought Index. Water Resour. Manag. 2019, 33, 3985–4006. [CrossRef] .

10. Malik, A.; Kumar, A. Meteorological drought prediction using heuristic approaches based on effective drought index: A case study in Uttarakhand. Arab. J. Geosci. 2020, 13, 1–17. [CrossRef] .

11. Malik, A.; Kumar, A.; Salih, S.Q.; Kim, S.; Kim, N.W.; Yaseen, Z.M.; Singh, V.P. Drought index prediction using advanced fuzzy logic model: Regional case study over Kumaon in India. PLoS ONE 2020, 15, e0233280. [CrossRef]

12. Malik, A.; Kumar, A.; Kisi, O. Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models. Comput. Electron. Agric. 2017, 143, 302–313. [CrossRef] .

13. Malik, A.; Kumar, A.; Kisi, O. Daily Pan Evaporation Estimation Using Heuristic Methods with Gamma Test. J. Irrig. Drain. Eng. 2018, 144, 04018023. [CrossRef] .

14. Malik, A.; Rai, P.; Heddam, S.; Kisi, O.; Sharafati, A.; Salih, S.Q.; Al-Ansari, N.; Yaseen, Z.M. Pan Evaporation Estimation in Uttarakhand and Uttar Pradesh States, India: Validity of an Integrative Data Intelligence Model. Atmosphere 2020, 11, 553. [CrossRef] .

15. Malik, A.; Kumar, A.; Kim, S.; Kashani, M.H.; Karimi, V.; Sharafati, A.; Ghorbani, M.A.; Al-Ansari, N.; Salih, S.Q.; Yaseen, Z.M.; et al. Modeling monthly pan evaporation process over the Indian central Himalayas: Application of multiple learning artificial intelligence model. Eng. Appl. Comput. Fluid Mech. 2020, 14, 323–338. [CrossRef]