Application of Artificial Neural Network and Genetic Algorithm Based Artificial Neural Network Models for River Flow Prediction

Arvind Yadav1*, Boggavarapu Bhanu Venkata Satya Vara Prasad2, Ramesh Kumar Mojjada3, Kiran Kumar Kothamasu2, Devendra Joshi3

1 Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India
2 Department of Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India
3 University Teaching Department, Chhattisgarh Swami Vivekanand Technical University, Bhilai 491107, Chhattisgarh, India

Keywords: artificial neural network, genetic algorithm, Mahanadi River, rainfall, water flow

https://doi.org/10.18280/ria.340608

Received: 7 October 2020
Accepted: 15 December 2020

1. INTRODUCTION

The water levels and water discharges are crucial components of hydrodynamics in river system. Rivers support land, marine life and river flow fluctuations have wide consequences for the water cycle, climate, crops and the ecosystem. The principal aspect of continental surface processes involving natural and manmade influences is the river transport of continental materials. A lot of work has been performed in the prediction and forecasting of the water discharges and level of water [1]. This research is important due to the river training, river management and power plant which are highly depended on flow rate of water in river basin system. The water flow variation is affected by human actions like man-made reservoirs, deforestation, construction of dams and urbanization.

Several models have been proposed to predict water flow discharge in a river, the characteristics of water flow are determined by climatic variables such as precipitation, temperature and humidity, etc. Rivers and compound channels have an effect on flooding and on the understanding of river morphology and, thus, over the past three decades, water flow discharge has gained a considerable amount of attention [2, 3]. The discharge corresponding to a certain depth is a well-known piece of work for hydraulic scientists, and because of its direct effect on flood risk mitigation. Thus, the accurate estimation of discharge for the overbank stage is very important. When the flow goes over-bank, the estimation becomes more difficult. For successful floodplain management, accurate prediction of flow at various locations of a river is very critical. Therefore, in flood forecasting that reduces flood harm, predicting water flow data with reasonable accuracy would be successful. In water resource operations such as surface water systems, the prediction of water flow discharge is also very important. The traditional predictions methods are not capable of the efficient prediction of the water flow due to complex nonlinear hydrological processes. Thus, Artificial Intelligence (AI) techniques are essential to prediction of water flow with higher accuracy by resolving these complex problems.

In past research, multiple linear regression (MLR), autoregressive (AR), multivariate autoregressive (MAR) and
other regression methods are used for prediction of water flow. However, these are basically linear models and have the ability to capture any linear relationship; however, they fail to model the presence of complex nonlinearity in water flow. To overcome these drawbacks, various AI techniques such as Artificial Neural Networks, genetic algorithm, fuzzy logic, vector support system, deep learning and different hybrid AI techniques have gained prominence in hydrological process modelling over the past decade. Using these AI approaches, the outflow from a barrage situated at the downstream end of a river can be predicted with more precision. Different researchers have been used the intelligent machine learning and soft computing techniques to handle the complex nonlinearity phenomenon in water resources and other various domains [4-17]. The ANN techniques are applied successfully in numerous interdisciplinary domain [15, 18-19].

Various artificial intelligence-based models like Genetic Algorithm (GA), Support Vector Machine (SVM) and Artificial Neural Network (ANN) have been developed in recent years in several major hydrological and other fields for the forecasting purpose [9, 11, 20-26].

One of the frequently used artificial intelligence-based technique is the artificial neural network (ANN) and have been successfully used by the various researchers for prediction of the water flow [27-29]. The ANN techniques are put in application successfully in numerous interdisciplinary domains [15, 18-19]. The ANN techniques have described the admissible outcome in rainfall forecasting, rainfall-runoff and sediment modelling [8, 30, 31]. Yadav [27] has been developed a forecasting model of water flow using ANN technique at Narag barrage in Mahanadi river, India.

Generally, the parameters for Artificial Neural Network method are selected by trial-and-error and/or grid search method to get a reasonably good result. However, this approach takes significantly large amount of computational time to come up with the parameters value, and not also guaranteed optimal or near-optimal solution of the problems. Different applications of artificial intelligence and machine learning models demonstrated the use of heuristic and metaheuristic approach for simultaneous optimization of associated parameters in artificial neural network model to overcome the limitation of trial-and-error methods. Several researches show that this approach is not only reduce the computational intensiveness, but also provide superior results. The main problems of the ANN are over- and under-fitting. Fitting problems are normally caused by poor selection of the ANN model parameters, such as the combination coefficient, and network topology, such as the hidden node size, i.e., nodes in hidden layers, number of hidden layers, initial weights, etc. The trial-and-error methods may not provide the optimum selection of these ANN’s parameters, and it is computationally expensive. Therefore, the selection of these parameters in the ANN models is an essential task in developing robust AI models.

The genetic algorithm is capable to overcome the problems of ANN’s model parameters selection which is provided global optimum solution of the model parameters [9, 25, 26]. So, in this study, genetic algorithm is used for ANN model parameters selection. The GA-ANN model has been successfully applied for the prediction and forecasting of streamflow, sediment, flood, rainfall and runoff [9, 30, 32-38].

On the basis of the potential of flood and capacity of the flood, the Mahanadi is ranked as the second biggest peninsular river. However, very little work on the prediction of water flow in the Mahanadi River has been carried out. As per author’s knowledge, there are no any studies available for water flow prediction by hybrid GA-ANN model with all model parameters optimisations simultaneously using water level, rainfall and temperature data in Mahanadi river basin, India. In this study, the main objective of the proposed work that to establish the robust novel GA-ANN model for prediction of the water flow using water level, rainfall and temperature parameters at Tikarapara (extreme downstream gauge station) gauge station in the Mahanadi river basin, India. The performances of this proposed GA-ANN model are compared with the ANN model. It is found that the hybrid GA-ANN model is provided better and satisfactory results than the single ANN model.

2. STUDY AREA DATA USED

Due to the amount of available data capturing the system's climate-hydrological relationship, the Mahanadi River basin, located in east-central India, was selected to illustrate the Artificial Intelligence Algorithms for estimation of water flow. There is a main branch length of 851 km in the Mahanadi basin (141,589 km²). It originates near Pharsiya village which lies in Raipur district of Chhattisgarh, India. The elevation of this originated place is almost 442 m above from the mean sea level. The Tikarapara gauge station in the Mahanadi River were used in the current study for monthly Precipitation, Water level, Temperature and Streamflow. The data observed covers a period of 1240 months (20 years), including data from January 1990 to December 2010. Meanwhile, the process of testing, training and validation for the development of applied models has been put in place. Normalization of the data is necessary so that the input range for each variable is within the interval (0, 1) [8, 9]. After gathering the data, normalization is carried out for each of the variables, i.e., rainfall, temperature, water discharge and water level which is briefly examined in Prabu et al. [26]. Before neural networks training, the data normalization is essential since input and output data are in different units and having different ranges. Data normalization is used to eliminate the dimension of the variables in order to maintain their uniqueness during the training in the models. Data normalization is also used to avoid any saturation effect that caused by the use of sigmoid activation function in ANN [34]. The location map of the Mahanadi river with Tikarapara gauge station is given in Figure 1.

![Figure 1. Map of the Mahanadi basin showing Hirakud Dam, main streams and Tikarapara locations](image-url)
3. METHODOLOGY

The one of the most popular AI techniques is ANN. Due to its limited data requirements, simplicity and cost-effectiveness, the ANN technique was found to be very suitable for the prediction and forecast modelling of a non-linear and dynamic system in water resources. The main advantage of the black-box ANN technique over traditional techniques is that it is not recognized the complex nature of the underlying processes which is explicitly described in mathematical form. The ANN is a class of statistical nonlinear models. The ANN has the ability to learn and handle the complex nonlinear relationships between variables from the actual measured data i.e., observed data. The ANN aims at developing learning technique through pattern recognition so that it can learn from data and estimate or forecast outputs [8]. The ANN has the capability to model any nonlinear complex mechanism which relates the water flow to hydro-geo-climatic data. The main objective of this analysis is to use temporal data (rainfall, temperature and water discharge) as inputs to assess the output of ANN for estimating water flow in the Mahanadi River.

As it adapts its own learning mechanism by evaluating the relationship between the input variables and output variables to achieve favorable outcomes. The ANN was recognized as one of the flexible models. In this analysis, ANN-based multi-layer perceptron (MLP) is used for water flow forecasting. The MLP is an artificial feed-forward neural network comprising an input layer, a hidden layer and an output layer. A particular number of neurons with an activation feature is included in each sheet. The number of input and output layer neurons must be equal to the input and goal parameters. By using the Levenberg Marquardt training algorithm, the MLP dependent ANN model is trained. The number of neurons in the hidden layers of the ANN is selected by the error statistics of the ANN model. The parameters of the ANN model are selected by trial-and-error method. The detailed description of MLP based ANN are given in Yadav et al. [8, 9].

The main goal is to demonstrate the fact that simultaneous selection of all ANN’s model parameters can significantly improve the robustness of the ANN model and performance of the model. The purpose of using GA is to select 4 major parameters of ANN models, namely - transfer function, number of neurons in hidden layer, combination coefficient, and connection and bias weights. All four neural network parameters are encoded to a binary string known as chromosome. The performance of GA depends on a high probability of crossover and a low probability of mutation. In this study uniform crossover with probability rate 0.6 was used. The low constant mutation probability was used in this study uniform crossover with probability rate 0.05. The maximum number of generation was used in this study is 50 as stopping criteria in the GA-ANN model. The population size in the GA-ANN model is 50. Our methodology is not unique, but its application in water flow prediction is unique. Moreover, the case study of Mahanadi River provides some insight into the application of artificial intelligence and machine learning methods in water flow prediction in river system.

4. RESULTS AND DISCUSSION

The MLP artificial neural network models which is based on the Levenberg Marquardt algorithm are proposed with single hidden layer through output, inputs and nodes(neurons). The value of $\mu$ is varied in the range of $0.001$ to $10 \times 10^9$ during this ANN training processes. The value of $\mu$ is increased or decreased by increasing and decreasing factors which contains the value $10$ and $0.1$, respectively. In order to reduce the complexity of structure and processing time in the artificial neural network, the size of the nodes in the hidden layer ranges between the $1$ and $32$. A trial and error procedure was used to select the number of neurons in the hidden layer and the learning parameters of the ANN model. In the ANN model, the hidden layer and output layer contains the tan-sigmoid transfer function and the pure linear activation function, respectively. The optimum number of iterations which is obtained during the ANN model's training is $16$. The optimum neurons in the ANN are $18$. Figure 2 demonstrates the difference in the performance of the ANN model during the training, testing and validation processes. Figure 3 provides the variance of training parameters with epochs during the training process. Figure 4 provides the configuration of the parameters of the artificial neural network during the training process.

![Figure 2. Performance variation with epochs in training, validation and testing phase](image1)

![Figure 3. Variation of training parameters with epochs during training phase](image2)
correlation coefficient (r) and error variance (VAR), have been used to test the efficiency of proposed models. The FFBP-LM model's error statistics (RMSE, r, MAE), MSE and error variance are provided in Table 1. The findings of the error statistics revealed that the ANN model had low RMSE (0.044041) and high r (0.9704) values. The r value between the performance value measured and expected is 0.9704, which is close to 1. From the results analysis, it is observed that the proposed ANN model is a very efficient and suitable model has been seen by the high value of r and low values of other error statistics. Moreover, there was also no underfitting and over-fitting in the model due to closed all error statistics values during validation, testing and training phase. A better generalization of the artificial neural network using the Levenberg - Marquardt algorithm was shown by the findings of the best-suited model.

![Figure 4. Structure of the Artificial Neural Network parameters during training phase](image)

**Figure 4.** Structure of the Artificial Neural Network parameters during training phase

![Figure 5. Comparison between observed and estimated sediment yield of ANN model in testing phase](image)

**Figure 5.** Comparison between observed and estimated sediment yield of ANN model in testing phase

![Figure 6. Scatter plot between observed and ANN estimated sediment yield based on training, validation and testing data](image)

**Figure 6.** Scatter plot between observed and ANN estimated sediment yield based on training, validation and testing data

![Figure 7. Generation wise fitness function profile during GA based neural network learning](image)

**Figure 7.** Generation wise fitness function profile during GA based neural network learning

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.044041</td>
</tr>
<tr>
<td>MAE</td>
<td>0.02008</td>
</tr>
<tr>
<td>r-correlation</td>
<td>0.97045</td>
</tr>
<tr>
<td>Error variance</td>
<td>0.00194</td>
</tr>
</tbody>
</table>

**Table 1.** Error statistics of ANN model at Tikarapara gauge station in Mahanadi River

It is also clear from the Figure 5 that estimated water flow values and observed water flow values are close to each other. Figure 6 shows the scatter plot between the predicted water flow and observed water flow using the ANN model. Most of the points are seen to be alongside the forty-five-degree line i.e., bisector line, where both observed and predicted values are roughly equal. Water flow is not overestimated or underestimated. A satisfactory and generalized result is
provided in the proposed model. All predicted water flow values by the ANN model are positive.

The variation of mean fitness and fitness value (RMSE) of the GA-ANN model during training phase corresponding to the generation is shown in Figure 7. The best fitness amongst all generations is found to be 0.00028 and its corresponding mean fitness value is 0.00131. The optimum value of optimized initial value of combinational coefficient (µ) in the Levenberg Marquardt algorithm 0.0008. The selected optimum neurons in the GA-ANN model is 32. Tan sigmoidal and pure linear sigmoidal functions are selected at hidden layer and output layer, respectively. The statistical analysis of errors which are calculated from the observed and model predicted suspended sediment yield is shown in Table 2. It was observed that root mean squared error (RMSE) is very low, and r is high. The mean absolute error (MAE) is also very low and consistent. Based on the RMSE, MAE, and r value, it reveals that this GA-ANN model has achieved good accuracy in prediction of suspended sediment yield.

![Figure 8](image8.png)

**Figure 8.** Comparison between observed and estimated sediment yield of GA-ANN model in testing phase

<table>
<thead>
<tr>
<th>Statistics</th>
<th>GA-ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.03017</td>
</tr>
<tr>
<td>MAE</td>
<td>0.01694</td>
</tr>
<tr>
<td>r-correlation</td>
<td>0.98616</td>
</tr>
<tr>
<td>Error variance</td>
<td>0.00088</td>
</tr>
<tr>
<td>MSE</td>
<td>0.00091</td>
</tr>
</tbody>
</table>

Table 2. Error statistics of GA-ANN model at Tikarapara gauge station in Mahanadi River

After the reliable GA-ANN model development, the performances of the best model are examined. The GA-ANN model represents the prediction model which is developed by the considering the rainfall. Temperature, water level and sediment yield, and applied on Tikarapara station. It is observed that GA-ANN model has smallest value of RMSE (0.03017) as comparative ANN model which having RMSE of 0.04404 in testing phase. On the basis of error statistics, it is observed from Table 1 and Table 2 that GA-ANN model provides better result than the ANN model of water flow at Tikarapara in Mahanadi river. The GA-ANN model has smaller value of the RMSE and MAE and greater value of r. It is also cleared from the graphical indicator that GA-ANN predicted water flow is and observed water flow is very closer as compare to the ANN model (Figures 5, 6, 8 and 9). Hence it reveals that GA-ANN model has better water flow prediction capability than the ANN model due to optimisation of all ANN’s parameters simultaneously using genetic algorithm. It indicates that the developed novel GA-ANN model has more generalisation capability to predict water flow in Mahanadi river basin. Thus, these approaches may be use for prediction of water flow where measurement of water flow in available.

5. CONCLUSIONS

The prediction of water flow is found using the ANN and novel GA-ANN model by considering water discharge, rainfall, sediment yield, temperature and water level data set as inputs at the Tikarapara gauge station in Mahanadi river basin, India. The parameters of these inputs have been found to be the most predominant parameter for water flow prediction. In previous literature, some artificial neural network models estimated negative values during low water flow, but in this study, all predicted GA-ANN and ANN model water flow values are positive. Moreover, the scatter plot and hydrograph between the predicted water flow by the model and observed water flow shows that the predicted water flows from the GA-ANN and the actual water flow are very closed. Therefore, the GA-ANN model performance is very good for...
prediction. It is also observed that the proposed GA-ANN model is provided satisfactory performance and it has more generalization capability as compare to the ANN model. This superiority of the GA-ANN model is due to the optimization of all ANN parameters simultaneously using the GA. This proposed model can be used efficiently for estimation of water flow where measurement of water flow is unavailable. In this study, the analysis only used data from a single gauge station and few temporal data sets. In future, more studies could be required using more data that requires spatial geological data from multiple gauge stations.

REFERENCES

https://doi.org/10.1504/IJAIP.2020.109512


https://doi.org/10.1016/j.jhydrol.2010.02.037


