ANN and GRNN-Based Coupled Model for Flood Inundation Mapping of the Punpun River Basin

Shashi Ranjan
Department of Civil Engineering, National Institute of Technology Patna, India
shashi.ce18@nitp.ac.in
(corresponding author)

Vivekanand Singh
Department of Civil Engineering, National Institute of Technology Patna, India
vsingh@nitp.ac.in

Received: 13 November 2022 | Revised: 22 November 2022 | Accepted: 25 November 2022

ABSTRACT

The Punpun River is primarily a rain-fed river. Forecasting rainfall accurately would enable an early evaluation of drought and flooding conditions. Therefore, having a flawless model for predicting rainfall is important for the hydrological analysis of any river basin. In this study, Artificial Neural Network (ANN)-based models were developed to predict rainfall and discharge in the basin. During the rainy season, water is spread in and around the area of the watershed, thus a General Regression Neural Network (GRNN)-based model was proposed for fast estimation of the inundation area during the flood taking as input cross-section, rainfall, and discharge. The proposed ANN-GRNN coupled model is the first of its kind for this study area. The assessment of the results shows that the proposed GRNN-based model is capable of estimating the water-spreading area.

Keywords-ANN; GRNN; rainfall; flood plain; Punpun river basin

I. INTRODUCTION

Rainfall is recognized as one of the most significant elements of the hydrological process [1]. If there is no evaporation, runoff, or infiltration, rainfall is the amount of water that falls on a flat surface over a specific period and is measured in millimeters (mm) above the horizontal surface. Forecasting rainfall accurately enables the early evaluation of drought and flooding conditions. Therefore, it is important to have a flawless rainfall prediction model. In a country like India, where most farmers rely on monsoons, good quality and quantity of water are very necessary for their crops. So, it is necessary to have advanced knowledge of the real amount of rainfall [2-5]. As several states in India are experiencing drought at the same time that many other states experience flooding, it is necessary to use an accurate and effective rainfall forecast model. This rainfall forecasting model could improve the handling of flood and drought problems. The ability of this model to anticipate rainfall in advance also provides sufficient time to plan transportation, lifesaving measures, and food and medication supply [6].

Time-series data are provided by scientific research, meteorological stations, GPS, sensor networks, etc. Time-series data have high dimensions and large volumes, and are continuously updated. Data analysis research is interested in the use of time series data for forecasting, pattern recognition, anomaly detection, pattern identification, clustering, classification, and segmentation [7]. Artificial Neural Networks (ANNs) are a well-established method for simulating complicated nonlinear and dynamic systems. ANNs are useful in creating an appropriate model, especially when the physical process relationship is unclear or the nature of the event exhibits chaotic qualities. Although prior knowledge of the system is necessary, ANNs lessen their reliance on this knowledge. As a result, there is no longer a requirement for a precise specification of the relationship's actual functional form, which the model aims to depict. Several studies utilized a variety of strategies to analyze time-series data, and among them, Generalized Regression, Pearson Coefficient, Fuzzy Inference System, Focused Time Daley Neural Network, and other Neural Networks (NNs) were found to perform sufficiently. Time series data have been analyzed in a variety of ways [8-10].

This study aimed to develop a rainfall prediction model for the Punpun river basin using ANN and GRNN-based models for fast estimation of the inundation area, taking as input cross-section, rainfall, and discharge. The three-layer Feed Forward Network (FFN) structure was used to construct the models and they were trained using back-propagation. The daily meteorological data of the Punpun river basin were considered for the development of the daily rainfall prediction models. The performance of the models was evaluated using various statistical indices.

II. STUDY AREA

Figure 1 shows a map of the Punpun river basin, which originates in the highlands of the Palamu district of the Indian state of Jharkhand at an elevation of 300 meters, 24°11' north and 84°9' east. Punpun merges with the river Ganga at Fatuha, 25Km downstream of Patna. It has no flow during dry months, but it has a significant discharge during monsoons which causes floods and submerges the nearby area. The water management of the basin is affected by both high discharge and no flow. From August to October, there is a lot of rain and the humidity of the region is moderate. The average annual rainfall ranges from 99 to 134cm, the highest at the upper levels (Palamu district). 80–87% of the yearly rainfall in the Punpun basin falls during the monsoon. The lower portion of the basin experiences uniform precipitations that do not change frequently. The Punpun river basin research area is 7,055Km^2. The region's geology is diverse, ranging from recent alluvium in the plains to granite, gneiss, and charnokites in the highlands.

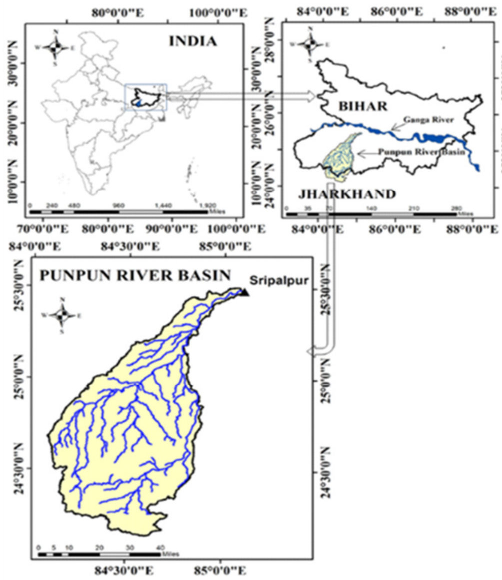

Fig. 1. The study area of the Punpun river basin.

III. METHODOLOGY

A. Data Collection

The daily observed rainfall data for 39 years (1 Jan 1982 to 31 December 2020) were collected from NASA's Agro Climatic Data (NASA POWER data access service (http://power.larc.nasa.gov/data-access-viewer/) for the Punpun river basin in India. Meteorological parameters, i.e., daily precipitation, relative humidity, wind speed, and maximum and minimum temperature, were used as input for the models to forecast daily rainfall. These data involved daily values of rainfall, temperature (min, max), relative humidity, wind speed, and solar radiation and were subjected to preanalysis and formulation of the database. 70% of the data were used for model calibration and the remaining 30% was used for validation.

B. Rainfall Prediction Modeling using ANN

This study used ANN models to predict daily rainfall in the Punpun River Basin, India. The combination of meteorological parameters as input for training ANNs was found as the most satisfactory for forecasting [11]. Therefore, the observed daily time series of temperature (T_{max}, T_{min}), relative humidity (*RH*), wind speed (*W*), and solar radiation (S_r) were considered. As the rainfall process is dynamic, current-day rainfall (P_{ij}) is considered to be a function of temperature, relative humidity, wind speed, and solar radiation. The functional form of rainfall modeling can be expressed as follows:

$$Pij = f(T_{max}, T_{min}, RH, W, S_r) \quad (1)$$

ANNs have been widely used to model basin hydrology and develop a non-linear relationship between the variables. ANNs have no fixed method for determining the pairs of input and output data. To fill this void, the number of data pairs used for training should be equal to or greater than the number of parameters (weights) in the network. This study used 14244 input-output datasets, separated into training from January 1982 to December 2014 and prediction/validation from Jan-2015 to Dec 2020. Rainfall prediction was modeled using the ANN tool in MATLAB. Figure 2 shows the approach to predicting daily rainfall using an ANN with LM training methods. One hidden layer was employed to mimic the rainfall in addition to the input and output layers. The model architecture used a feedforward neural network and backpropagation and was made up of several neurons dispersed across and related to each other. At first, the network was trained on a set of paired data to manage the input-output meaning. Then the weights of the connections between neurons were fixed, and the network was used to determine the classification of new data.

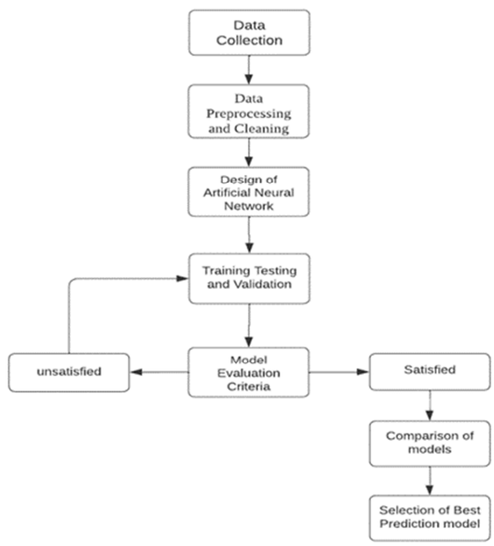

Fig. 2. Method for the prediction of daily rainfall in Punpun river basin.

This study used the most suitable algorithm based on the model evaluation criteria for forecasting, the feed-forward method with back-propagation [12]. The input data for the system were the daily meteorological data of the Punpun river basin. The training process used 70% of the time-series data, while each of the testing and validation processes used 15%. The purpose of the training was to identify the connections, weights, and biases that allow the neural network to estimate outputs that are reasonably similar to the measured outputs. The training datasets were also used to reduce errors. By changing the weights and biases in the ANN, different training methods can be used to train the input parameters to produce an output with a smaller global error [13]. The Levenberg-Marquardt (LM) [14] training algorithm was used in the training process of the developed ANNs. Table I shows the training parameters.

TABLE I. DETAILS OF TRAINING PARAMETERS

Parameters used for training ANN	
Network type	Feed Forward Neural Network with Back-Propagation
Training function	Levenberg Marquardt (LM)
Adoption learning function	Learn GDM
Performance function	Mean Square Error (MSE); Regression (r)
Transfer function	Hidden layer-Tansigmoid; Output layer- Linear
No. of neurons used in the hidden layer	5-45

C. Levenberg-Marquardt (LM) Training Algorithm

The LM training algorithm is a standard and widely applied iterative method that, in most cases, deals with curve-fitting arrangements and trains ANNs much more quickly than the typical back-propagation algorithm. The LM method finds the base of a multivariable function and expresses it as the sum of squares of nonlinear genuine valued functions. This approach combines the advantages of two strategies by switching the parameter updates between the Gauss-Newton and the gradient descent update. The LM algorithm alters the conventional Newton algorithm to determine the best answer to a minimization issue, as:

$$X_{n+1} = X_n - \{J^T J + \mu I\}^{-1} J^T e \quad (2)$$

where x is the weight of NNs, J is the Jacobian matrix, e is the residual error vector, and μ represents the scaler that controls the learning process. As the LM algorithm is computationally and memory intensive, it is best suited for small networks.

D. General Regression Neural Network (GRNN)

The General Regression Neural Network (GRNN) model was used to estimate the river water spread by using rainfall and discharge as estimated by the ANN model and river-cross section. GRNN is a neural network-based function predicting algorithm, more desirable than other neural network models as it doesn't require any data for the iterative training process [15, 16]. GRNN is based on nonlinear regression and can utilize the training data to approximate any arbitrary continuous nonlinear function. For GRNN, a weighted average of the training dataset's outputs is used to calculate the new output. The Euclidean distance between a given pattern and the training dataset is used to estimate the weight of that pattern. A smaller distance results in a higher weight for the pattern while a larger distance results in a lower. A GRNN model has four fundamental layers: input, pattern, summation, and output [17]. Figure 3 shows the GRNN network model. The input layer contains all input data from the simulation. The pattern layer determines the relevant weight and computes the Euclidean distance. The inputs' Euclidean distances are used as the basis for determining how much weight to give to each pattern. Equation 3 calculates the new output based on the new input and training datasets.

$$Y(X) = \frac{\sum Y_i e^{-\left(\frac{d_i^2}{2\sigma^2}\right)}}{\sum e^{-\left(\frac{d_i^2}{2\sigma^2}\right)}} \quad (3)$$

where d_i is the Euclidean distance or spread between the new (X) and the training input (X_i), the higher the value of, the less likely it is that the output will be close to the inputs' extreme values and vice versa. When the distance d_i value is small, the weight term returns a relatively large value and vice versa. If d_i is zero, the weight term gives a result of 1, meaning that the output of the test data will be the same as the output of the training sample.

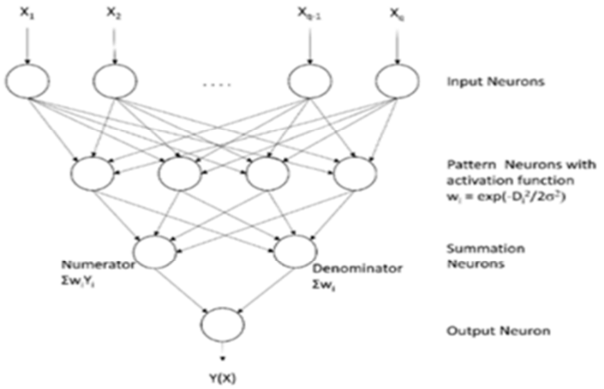

Fig. 3. The basic network diagram of a GRNN model.

The spread value σ is the only parameter that needs to be estimated for the GRNN model. The training process involves determining the ideal value of σ. The best method to determine the spread value is to use an optimization algorithm to reduce Mean Squared Error (*MSE*). The entire dataset is split into two parts, training and testing, to obtain the ideal value of the spread. The best value of the spread is then obtained by minimizing *MSE* after GRNN is applied to the test data based on the training data. As stated previously, the model is created to estimate the extent of water spreading, particularly during the monsoon. A possible representation of an input pattern (k) is as follows:

$$X_k = [Q_k^t, Q_k^{t-1}, Q_k^{t-2}, S_k^t, S_k^{t-1}, S_k^{t-2}, P_k^t, P_k^{t-1}, P_k^{t-2}] \quad (4)$$

where Q represents the discharge, P is precipitation, and S is the river cross-section. The corresponding output pattern is given by:

$$Y_k = [SW_k^{t+1}] \quad (5)$$

Figure 4 shows the GRNN model applied in this study. Any arbitrary pattern k for the Punpun river basin, where recorded water level data is available, is represented by the input pattern X_k^i. The pattern layers calculate the weight (w_n^i) of each individual pattern as well as the distance (d_n^i) between the input and training patterns. There are N different training pattern options. N neurons will therefore be present in the pattern layer. The weight of the calculated pattern is multiplied by the corresponding output in the summation layer, the result is summed at the numerator neuron, and all the weights are added up at the denominator neuron. The output neuron uses (3) to calculate the output for the arbitrary vector. The spread value σ, which is the only parameter in the model, can be calculated using an optimization algorithm. The model can now be used to estimate the extent of water spreading after it has been trained or after the optimal spread value has been determined. This model was constructed for the i^{th} cross-section of the watershed along the Punpun river line. Thus, the model was named $GRNN_i$.

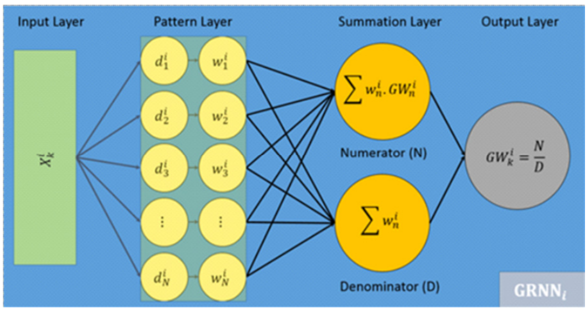

Fig. 4. GRNN architecture of the model.

E. *Model Performance Evaluation*

The prediction performance of each model was determined by the statistical and hydrological indices and quantitative comparison between the observed and predicted values.

1) *Statistical Indices*

a) *Mean Square Error (MSE)*

MSE is used to compare differences between observed and predicted values. *MSE* is zero for a perfect fit, and increased values indicate a higher deviation between predicted and observed values. *MSE* is determined using:

$$MSE = \frac{1}{n}\sum_i^n \left(P_{observed} - P_{predicted}\right)^2 \quad (6)$$

b) *Correlation Coefficient*

The correlation coefficient (r) is an indicator of the degree of closeness between the observed and predicted values and provides the level of explained variance between them [18]. The correlation coefficient is given by:

$$r = \frac{\sum_i^n (P_{observed} - \underline{P}_{observed})(P_{predicted} - \underline{P}_{Predicted})}{\sqrt{(P_{observed} - \underline{P}_{predicted})^2}\sqrt{(P_{predicted} - \underline{P}_{Predicted})^2}} \quad (7)$$

2) *Nash-Sutcliffe Efficiency (NSE) Hydrological Indice*

The NSE was developed by Nash and Sutcliffe in 1970. It is also called the coefficient of efficiency. It provides the proportions of variance of the observation for a model and is widely used in hydrology [19]. NSE has a range between -∞ to 1, and a value of 1 indicates a perfect match. It is determined by:

$$NSE = 1 - \frac{\sum_{i=1}^n (P_{observed} - P_{predicted})^2}{\sum_{i=1}^n (P_{observed} - \underline{P}_{observed})^2} * 100 \quad (8)$$

where $P_{observed}$, $P_{predicted}$, and $\underline{P}_{observed}$ are observed, predicted, and mean rainfall values at the Punpun River basin, and n is the number of observed datasets.

IV. RESULTS AND DISCUSSION

The performance of the models was qualitatively and quantitatively evaluated by visual observation and various statistical and hydrological indices.

A. *ANN-Based Models for Rainfall Prediction during Training and Testing*

The model was trained by varying the number of neurons in the hidden layer. Each NN was identified by the number of input, hidden, and output nodes. Table II shows the performance evaluation indices for rainfall estimation during training and testing. The networks having lower *MSE* and higher r and *NSE* values were chosen as the best prediction models. Table II shows that the performance of model 5-15-1 was better during training and testing. The r values for the ANN networks ranged from 0.76 to 0.93 and from 0.82 to 0.97 during training and testing, respectively. *MSE* values ranged from 7.08 to 20.7 and from 3.35 to 16.7 during training and testing, respectively.

TABLE II. PERFORMANCE OF VARIOUS ANN MODELS

Training				Testing			
Network	**MSE**	**r**	**NSE**	**Network**	**MSE**	**r**	**NSE**
5-5-1	12.45	0.83	81	5-15-1	10.45	0.87	88
5-10-1	16.78	0.81	78	5-19-1	13.78	0.85	86
5-15-1	**7.08**	**0.93**	**86**	**5-15-1**	**3.35**	**0.97**	**94**
5-20-1	10.6	0.85	82	5-20-1	8.6	0.91	9
5-25-1	8.47	0.9	83	5-25-1	5.47	0.95	92
5-30-1	13.48	0.82	8	5-30-1	11.48	0.86	87
5-35-1	19.48	0.77	73	5-35-1	15.48	0.83	85
5-40-1	17.33	0.8	77	5-40-1	14.33	0.84	86
5-45-1	20.7	0.76	7	5-45-1	16.7	0.82	84

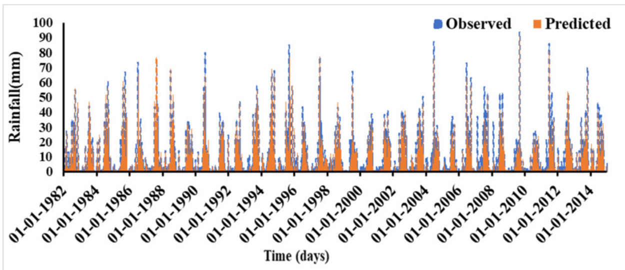

Fig. 5. Observed and predicted rainfall during training period.

Figure 5 shows the observed and predicted values of rainfall during the training period (1982-2014) using the 5-15-1 ANN model, while Figure 6 shows the same values for the testing period (2015-2020). These figures show that there is close or nearly close agreement between the observed and predicted rainfall, and the overall shape of the predicted rainfall curve is similar to the observed rainfall. The quantitative evaluation of the model, based on NSE, shows a significant relationship between observed and predicted rainfall. The NSE values for the 5-15-1 ANN were 84% and 94% during training and testing, respectively. Thus, the 5-15-1 ANN was selected as the best for rainfall prediction. Figure 7 shows the correlation between the observed and predicted rainfall for the 5-15-1 ANN during training and testing.

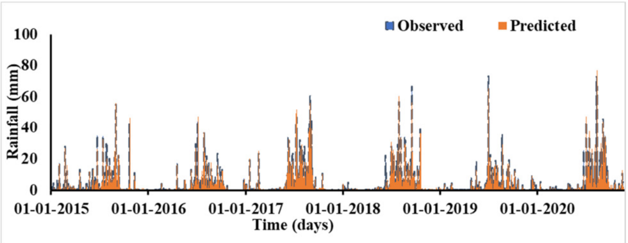

Fig. 6. Observed vs predicted rainfall for the 5-15-1 ANN model during the testing period.

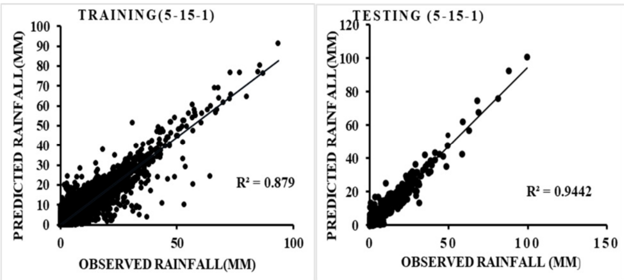

Fig. 7. Correlation of observed and predicted rainfall for the 5-15-1 ANN during training (left) and testing (right).

The quantitative and qualitative performance during training and testing was found satisfactory, as these models predicted rainfall with acceptable accuracy. Based on qualitative and quantitative performance evaluation, the ANN models gave satisfactory results for the study area.

B. GRNN Model for Punpun River Flood Mapping

A scattered plot was plotted for the surface water spread between the simulated and the actual data at different cross-sections, and a map was created in ArcGIS for better visualization [20]. Table III and Figure 8 show that the variation of the water spread in form of fluctuation is satisfactory and the R^2 value varied from 0.69 to 0.91.

TABLE III. CORRELATION OF ACTUAL AND PREDICTED WATER SPREAD

Cross section name	R^2 value b/w actual and GRNN predicted
1	0.696
2	0.769
3	0.672
4	0.619
5	0.711
6	0.699
7	0.912
8	0.811
9	0.678
10	0.745

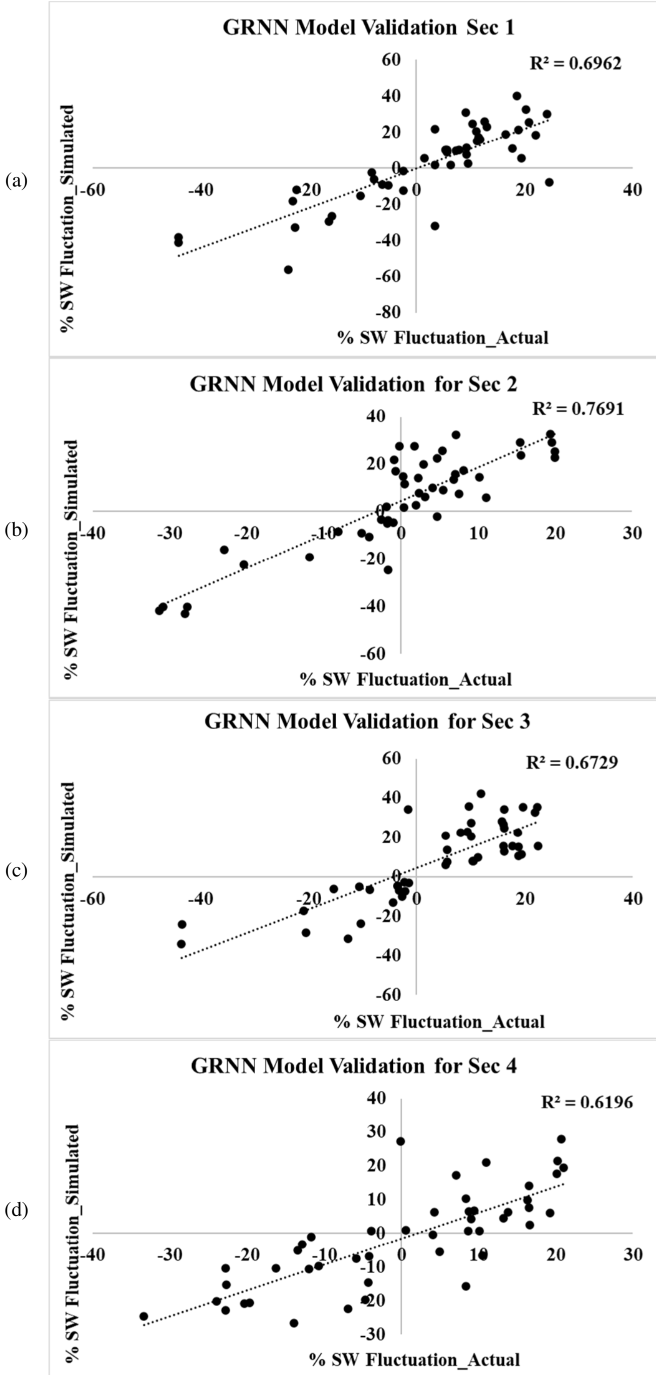

Fig. 8. Scatterplots between actual and predicted spread water fluctuation by the GRNN model of section 1, 2, 3, and 4 cross-sections, respectively.

Figure 9 shows a water spreading map, where there is not much water inundation in the Punpun river basin. The map shows the width of the river to the water spread area of that river section. The main reason behind this is that Punpun is a rainfed river. At some cross sections, mainly in the middle part of the river basin, the water is spreading away from the main streamline to the flood plain.

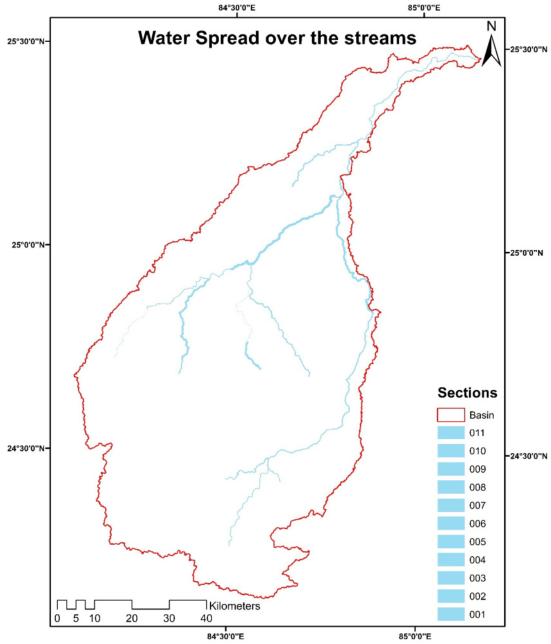

Fig. 9. Actual spread water fluctuation along the river cross-sections.

V. CONCLUSIONS

This study showed that feedforward backpropagation networks using the LM training algorithm produced better results compared to others. Rainfall is more influenced by factors such as minimum and maximum temperature, relative humidity, wind speed, and solar radiation, so these characteristics cannot be ignored in predicting. Instead of pre- and post-monsoon months, mid-monsoon months provide more accurate results for predicting rainfall. The initial flooded area map was created using GRNN, a popular machine-learning method with strong learning performance. The initial flood map projected by the GRNN was subjected to easy-to-implement two-stage post-processing to further reduce missed and false water spread detections and increase the accuracy and reliability of flood mapping [20]. Therefore, the proposed GRNN model was capable of producing the water spread fluctuation over the cross-section. The limitation of the coupled model was that this study area is mainly a rainfed river system and the coupled model can be used in other river basin systems such as the Himalayan River system to get more satisfactory results.

REFERENCES

[1] S. Rana, V. Deoli, and S. R. Chavan, "Detection of abrupt change in trends of rainfall and rainy day's pattern of Uttarakhand," *Arabian Journal of Geosciences*, vol. 15, no. 7, Mar. 2022, Art. no. 618, https://doi.org/10.1007/s12517-022-09883-w.

[2] V. Deoli and S. Rana, "Seasonal trend analysis in rainfall and temperature for Udaipur district of Rajasthan.," *Current World Environment*, vol. 14, no. 2, pp. 312–319, 2019.

[3] A. Towheed and R. Thendiyath, "Spatiotemporal Rainfall Dynamics in Kosi Basin Using Wavelet Analysis," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7578–7584, Oct. 2021, https://doi.org/10.48084/etasr.4372.

[4] A. Towheed and T. Roshni, "Linking climate change to soil loss estimation in the Kosi river basin, India," *Journal of Water and Climate Change*, vol. 12, no. 6, pp. 2338–2363, Mar. 2021, https://doi.org/10.2166/wcc.2021.259.

[5] K. Praveen and L. B. Roy, "Assessment of Groundwater Quality Using Water Quality Indices: A Case Study of Paliganj Distributary, Bihar, India," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8199–8203, Feb. 2022, https://doi.org/10.48084/etasr.4696.

[6] A. Rawat, P. Kumar, and V. Deoli, "Daily Monsoon Rainfall Prediction using Artificial Neural Network (ANN) for Parbhani District of Maharashtra," *International Journal of Current Microbiology and Applied Sciences*, vol. 8, no. 12, pp. 1949–1963, Dec. 2019, https://doi.org/10.20546/ijcmas.2019.812.233.

[7] P. Esling and C. Agon, "Time-series data mining," *ACM Computing Surveys*, vol. 45, no. 1, Sep. 2012, Art. no. 12, https://doi.org/10.1145/2379776.2379788.

[8] Y. Dash, S. K. Mishra, and B. K. Panigrahi, "Rainfall prediction for the Kerala state of India using artificial intelligence approaches," *Computers & Electrical Engineering*, vol. 70, pp. 66–73, Aug. 2018, https://doi.org/10.1016/j.compeleceng.2018.06.004.

[9] A. Raja and G. T, "Drought Prediction and Validation for Desert Region using Machine Learning Methods," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 47–53, 2022, https://doi.org/10.14569/IJACSA.2022.0130707.

[10] C. Sireesha, T. Roshni, and M. K. Jha, "Insight into the precipitation behavior of gridded precipitation data in the Sina basin," *Environmental Monitoring and Assessment*, vol. 192, no. 11, Oct. 2020, Art. no. 729, https://doi.org/10.1007/s10661-020-08687-3.

[11] N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi, "An artificial neural network model for rainfall forecasting in Bangkok, Thailand," *Hydrology and Earth System Sciences*, vol. 13, no. 8, pp. 1413–1425, Aug. 2009, https://doi.org/10.5194/hess-13-1413-2009.

[12] L. Zhao, F. E. Hicks, and A. R. Fayek, "Applicability of multilayer feed-forward neural networks to model the onset of river breakup," *Cold Regions Science and Technology*, vol. 70, pp. 32–42, Jan. 2012, https://doi.org/10.1016/j.coldregions.2011.08.011.

[13] A. Ehret, D. Hochstuhl, D. Gianola, and G. Thaller, "Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle," *Genetics Selection Evolution*, vol. 47, no. 1, Mar. 2015, Art. no. 22, https://doi.org/10.1186/s12711-015-0097-5.

[14] D. Yadav, R. Naresh, and V. Sharma, "Stream flow forecasting using Levenberg-Marquardt algorithm approach," *International Journal of Water Resources and Environmental Engineering*, vol. 3, no. 1, pp. 30–40, Jan. 2011.

[15] D. Kumar and R. K. Bhattacharjya, "GRNN Model for prediction of groundwater fluctuation in the state of Uttarakhand of India using GRACE data under limited bore well data," *Journal of Hydroinformatics*, vol. 23, no. 3, pp. 567–588, Mar. 2021, https://doi.org/10.2166/hydro.2021.108.

[16] T. Roshni, M. K. Jha, R. C. Deo, and A. Vandana, "Development and Evaluation of Hybrid Artificial Neural Network Architectures for Modeling Spatio-Temporal Groundwater Fluctuations in a Complex Aquifer System," *Water Resources Management*, vol. 33, no. 7, pp. 2381–2397, May 2019, https://doi.org/10.1007/s11269-019-02253-4.

[17] S. A. Hannan, R. R. Manza, and R. J. Ramteke, "Generalized Regression Neural Network and Radial Basis Function for Heart Disease Diagnosis," *International Journal of Computer Applications*, vol. 7, no. 13, pp. 7–13, Oct. 2010, https://doi.org/10.5120/1325-1799.

[18] K. R. K. Reddy and V. Singh, "Numerical Model to Simulate Soil Wetting Pattern under Drip Irrigation System," *Agricultural Science Digest*, Aug. 2022, https://doi.org/10.18805/ag.D-5619.

[19] S. Ranjan and V. Singh, "HEC-HMS based rainfall-runoff model for Punpun river basin," *Water Practice and Technology*, vol. 17, no. 5, pp. 986–1001, Apr. 2022, https://doi.org/10.2166/wpt.2022.033.

[20] Q. Zhang, P. Zhang, and X. Hu, "Unsupervised GRNN flood mapping approach combined with uncertainty analysis using bi-temporal Sentinel-2 MSI imageries," *International Journal of Digital Earth*, vol. 14, no. 11, pp. 1561–1581, Nov. 2021, https://doi.org/10.1080/17538947.2021.1953160.