xIlkogretim Online - Elementary Education Online,2021; Vol 20 (Issue 5): pp.3008-3021
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.05.326

Forecasting of Flood in Upper Yamuna Basin by Using Artificial Neural Network and Geoinformatics Techniques& Learning

MS. Parveen Bano, Research Scholar, Department of computer Science, Jagannath University, NCR, Bahadurgarh, Haryana.
Dr. Rajender Singh,Department of Geography, Shivaji College, University of Delhi.
Dr. Gaurav Aggarwal, Head of the Department, Department of computer Science, Jagannath University, NCR.

ABSTRACT- River flow forecasting is required to supply basic information on a wide range of problems associated with the planning and operation of river systems. While traditional models are important for understanding hydrological processes, still accurate predictions are required at some specific locations. In this context ANN (Artificial Neural Network) “black box” in nature, is emerging as most successful machine learning techniques with flexible mathematical structure, capable of showing non-linear association between input and output data. Therefore,in current years, ANN models are used frequently for forecasting of flood. Most of the previous work of ANN techniques in forecasting has been done by taking rainfall as input and runoff data as output. These forecasting has not considered other flood causative factors. Therefore, the main goal of this study is to develop a flood model by considering many factors which are responsible for occurrence of flood. Also, it has been observed that space and air-based observations of earth gives a unique way to monitor and assess the floods. So, this study has tried to take an advantage of both ANN and GIS by integrating these techniques together. Therefore, ANN and GIS are used for modeling and simulations of flood prone area in Yamuna Nagar district of Haryana. The model was implemented in MATLAB. Maps related to flood causative factors such as rainfall, slope, elevation, flow accumulation, soil, land use and geology are prepared with the help of GIS, Remote Sensing data and field surveys. Water level is produced by ANN model and then flood map showing flow accumulation is generated using GIS.
The model performance is measured in terms of coefficient of determination (R^2), sum squared error(SSE), the mean squared error(MSE) and the root mean squared error(RMSE).

Keywords: **Artificial Neural Network (ANN), Geographic Information System (AAN), MATLAB, Coefficient of determination (R2), Sum Squared Error (SSE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).**

I. INTRODUCTION

Flood forecasting can be defined as a process of estimating and predicting the magnitude, timing and duration of flooding based on known characteristics of a river basin, with the aim to prevent damages to human life, to properties, and to the environment

Floods are a recurring phenomenon in India. Due to different climatic and rainfall patterns in different regions, some areas experience catastrophic flooding and another area experiences simultaneous drought. It is considered one of the most destructive natural phenomenon that can cause damage to both life and property every year. It is therefore has become one of major concern in world over. Forecasting river flow after heavy rain is important for public safety, environmental issue and water management.

The current studies revealed that many part of the state of Haryana are also prone to flooding. The devastating floods hit Haryana many times. In 1977, 1978, 1980,1983,1988,1993 and 1995, 1996 floods occurred in Haryana. Floods have been causing extensive damage not only to standing crops but also loss of lives and cattle. The floods in Haryana can occur because of some natural reasons such as its physiographic situation which makes a depressional saucer shape zone around the Delhi-Rohtak-Hisar-Sirsa axis and it has a poor natural drainage system and sometimes the heavy precipitation becomes a major contributing factor in causing flood as such in case of Rohtak flood, 1995. The state receives anaverage rainfall of about 650 mm. The average annual rainfall varies from less than 300mm in the western and south western parts of Sirsa, Hisar and Bhiwani districts along the Rajasthan border to over 1100mm in the north-eastern Shivalik hilly tracts of Panchkula and Yamuna Nagardistricts along Himachal Pradesh border The problem of floods is further accentuated by theexistence of human-made barriers like the networks of roads and canals, which obstruct the natural flow of water and sometimes Drainage systems back up because they cannot cope up with the volume of water or are blocked by rubbish and garbage. On the other hand indiscriminate use of water for irrigation and development of low-lying areas and depressions has also created problem of drainage congestion and water logging which create a havoc of flooding. According to assessment of Rashtriya Barh Ayog and as reported by states to the 11th plan working group, flood prone area in Haryana is 23.50 lakh hectares. In flood manual of Haryana, there are 102 vulnerable points in Haryana which need special attention during monsoon (*HSDMA report*).

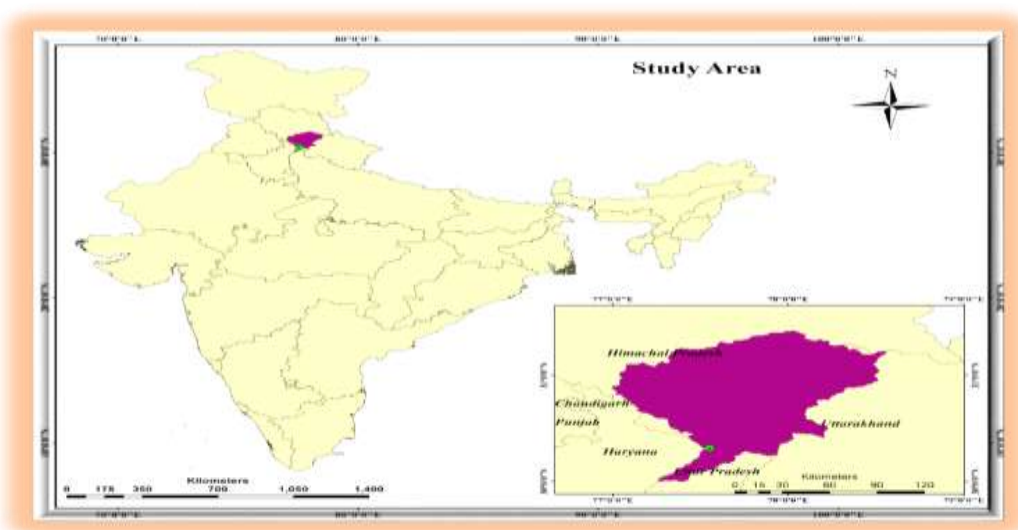
For this purpose, many tools and techniques are used to solve the problems of flood inundation. In recent years, some of the approaches such as ANN, Remote Sensing, and GIS are increasingly used for flood modelling. To increase the accuracy of flood model and to handle some of the limitations of hydrological model many other new techniques such as ANN, Fuzzy logic, Neuro-fuzzy are used to predict flood. All these models are predicting flood by taking rainfall as input data and runoff as output data. Therefore the aim of this study is to propose a flood model using various flood causative factors and integrate them with ANN and GIS to make and simulate flood prone area at upper Yamuna River Basin, Haryana.

There are many features in ANN which make this technique valuable and attractive for forecasting task. These are:-

- ANNs are data driven self-adaptive method which learn from past and capture functional relationship among input and output if that is hard to describe.
- ANNs are generalize in nature which means they can correctly infer the unseen part if the data is noisy.
- ANNs are known as universal functional approximators' means it can approximate any continuous function to any desired accuracy.
- ANNs are nonlinear so more general and flexible approach for forecasting.

Study Area

Yamuna Nagar district in Haryana is located between 29º 55 '30 º 31 'north latitude and 77º 00' 77º 35 'east longitude. The total geographical area of the district is 1756 sq km, which is 4% of the total area of the state. District boundary: Himachal Pradesh in the north, Uttar Pradesh in the east, Ambala district in the west, Colonel and Kurukshetra districts in the south. Yamuna Nagar is the second largest industrial city of Haryana. Population explosion, uncontrolled urbanization and industrialization have led to high waste production in the Yamuna. Industrial waste contains significant amounts of volatiles, biochemical and their derivatives. Most industries are small, have no sewer lines, and discharge industrial waste into offline channels and streams, polluting the air, water and soil. (Sharma et al., 2013)



Location map of study area

II. DATABASE AND METHODOLOGY

In thisstudy, to forecast flood, seven inputs are taken these are as follows:-

1. Slope
2. Digital Elevation Model
3. Soil
4. Flow Accumulation
5. Geology
6. Land use/ Land cover
7. Rainfall

Slope

Slope map indicates the topography of a surface. The gradient map of the study area is prepared using ASTER DEM data.The general slope of the district is from east-east to south-west and the major rivers of the district are Yamuna, Saraswati, Markanda and Chautang. A part of the semi-mountainous region, the district is divided into several nallas and rivers. All streams except the Yamuna are seasonal. The Yamuna is the source of the Western Yamuna canal system which the State of Haryana needs for irrigation. Streams / rivers flowing through the lower urban hills cause heavy flooding during monsoons. This required extensive planning of the drainage system.

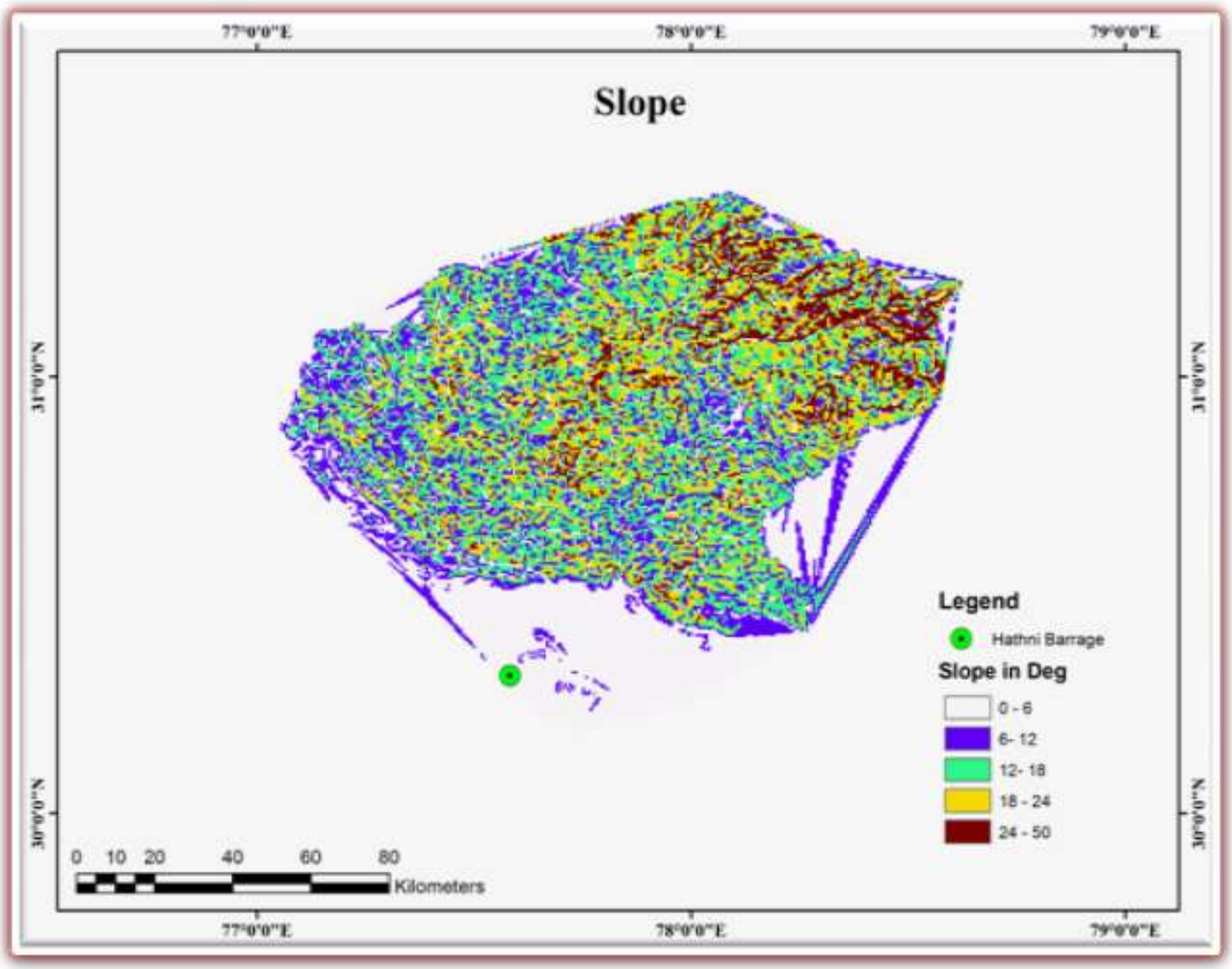


Figure: Slope Map

Digital Elevation Model

Thecurrent research activities are provided in and around the Yamuna River Hathnikund Barrage, the size of the study area extracted using DEM (Digital Elevation Model) data. Area snap PowerPoint tool is used. The Snap PowerPoint tool is used to confirm the selection of high accrued flow points when defining drainage beds using a watershed device. Move the snap power point searches and spreading area to that location at a second distance around the specified pulse points for the cell with the maximum collected flow. If the input point data is a point feature class, it will be transformed into a local raster for processing. When there is only one input pole point location, the output size is the size of the cache raster. If there is more than one pole point location, theenvironment output is determined by environmental sites.

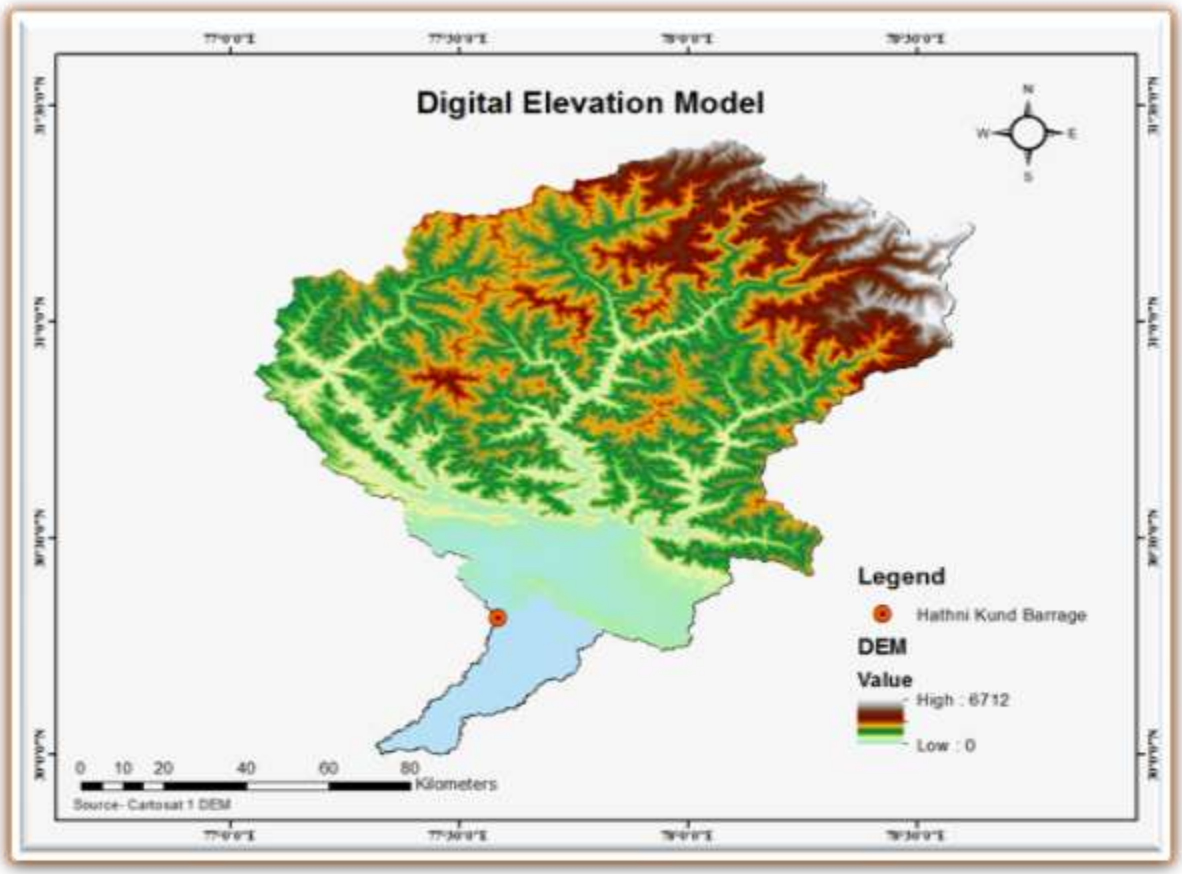


Figure: DEM Map

Soil

Soil is an important part of the study of a river basin.Soil classification is based upon the soil survey interpretation. Based on interpretation the potentialities and limitation of the soil can be obtained and such information are used to construct database using GIS. In upper Yamuna basin our study have been covered with six types of soils. These are Bhabar, Mountain meadow, Red, Brown, Yellow, SubMountain and older alluvial soils.

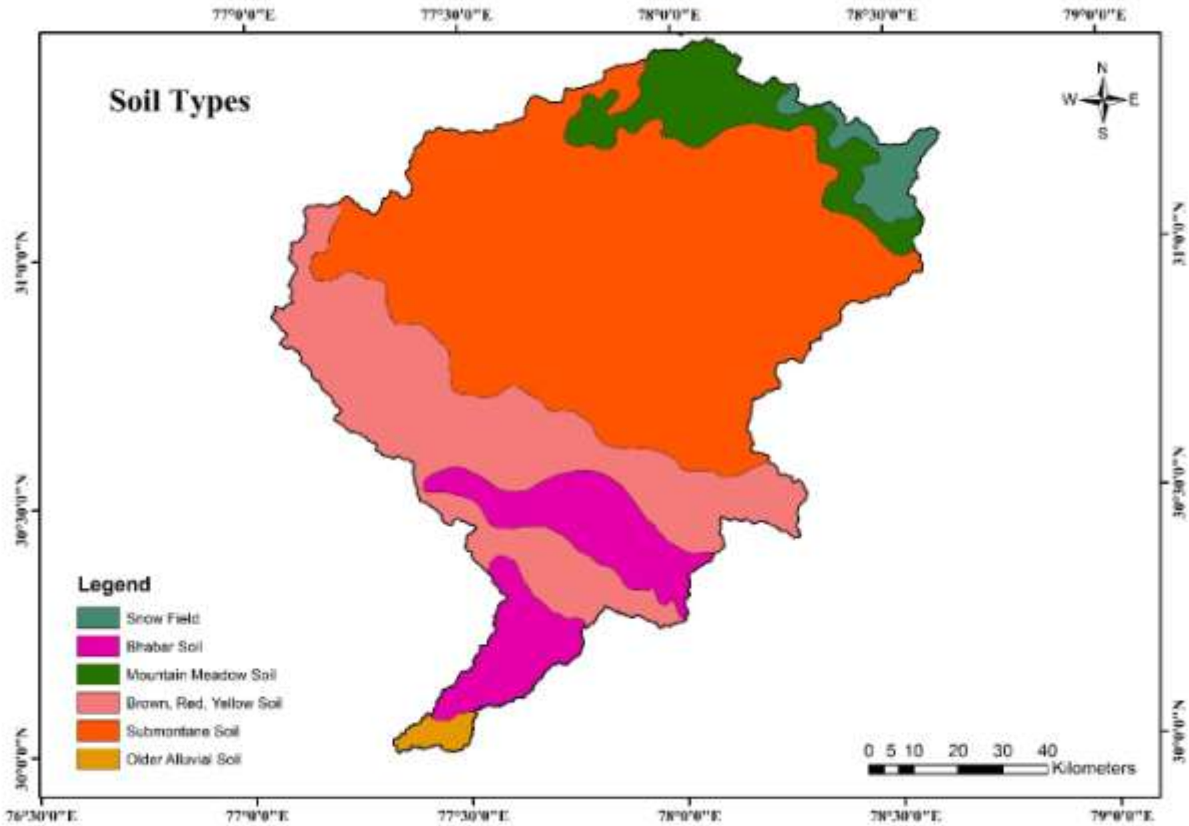


Figure: Soil Map

Flow Accumulation

Flow accumulation data was calculated in Arc GIS 10.5, The Flow Accumulation tool calculates accumulated flow as the accumulated weight of all cells flowing into each downslope cell in the output raster. If no weight raster is provided, a weight of 1 is applied to each cell, and the value of cells in the output raster is the number of cells that flow into each cell.

Cells with a high flow accumulation are areas of concentrated flow and may be used to identify stream channels. This is discussed in Identifying stream networks. Cells with a flow accumulation of 0 are local topographic highs and may be used to identify ridges.

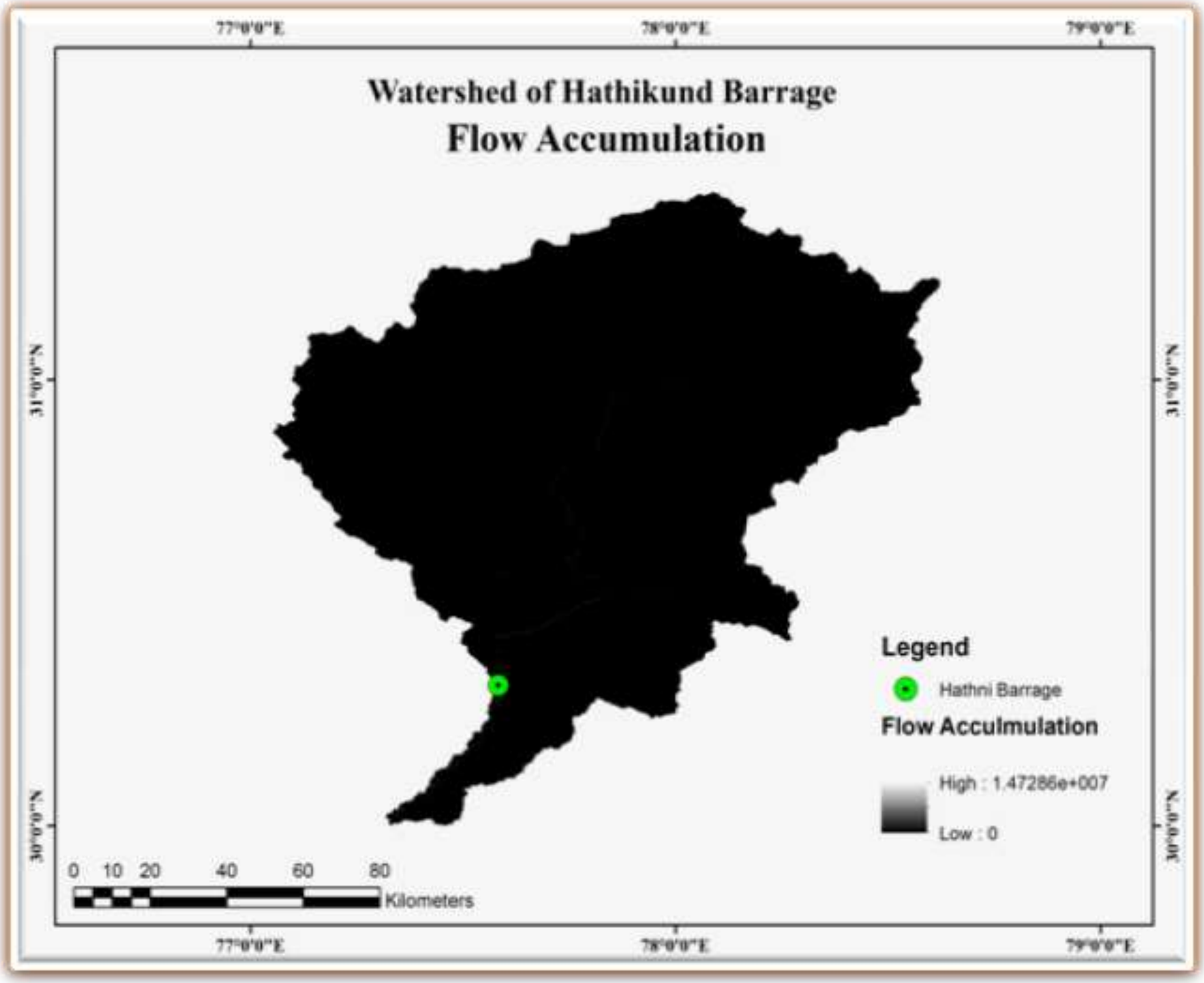


Figure: Flow accumulation map

Geology

Four main geomorphological areas can be identified in the catchment: (a) Archaean (b) Cambrian-Ordovician in the middle part of the basin; (c) Quaternary which occupy the south western part of the area; and (d) Palaeozoic lower part of the catchment area.

The lithology of a rock unit is a description of its physical characteristics visible at outcrop, in hand or core samples, or with low magnification microscopy. The geomorphologic map of the entire project area has been prepared mainly through Liss III satellite data on 1:50,000 scale. Some photographic details have been transferred from Toposheet to the base map along with the interpreted units through satellite data.

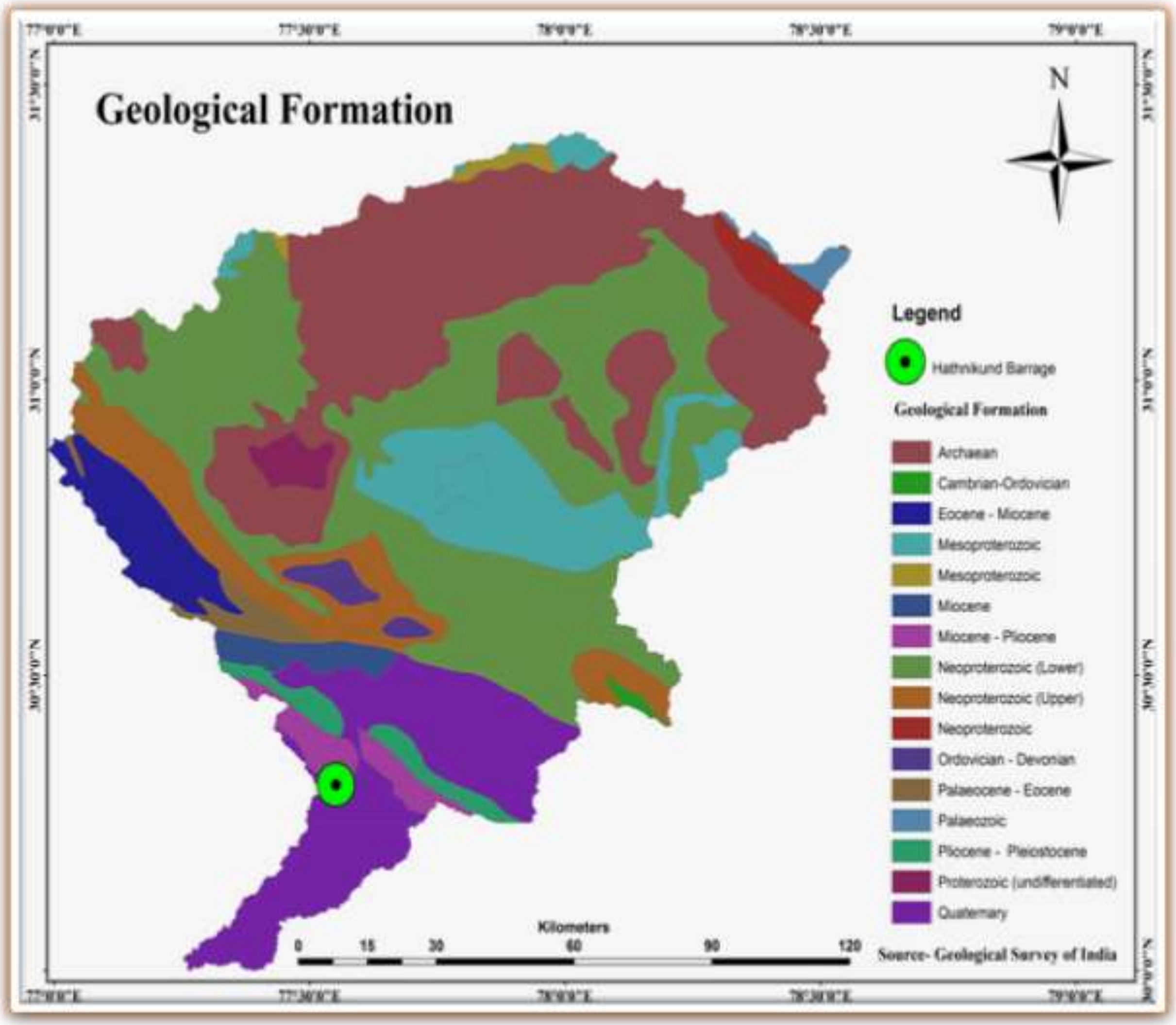


Figure: Geology Map

Land use/ Land cover

The main aim of the present research work is to monitor the land use/ land cover of the present study area and to check the kind of impact on the flood. For this purpose Landsat 8 image was downloaded from USGS Earth explorer. These processes include impact of climate, geologic and topographic conditions on the distribution of soils, vegetation and occurrence of water. For better development and Flood management, it is necessary to have timely and reliable information of land use and land cover. Keeping the above views in mind, Prepared a land use/land cover map using Landsat 8 data.LU / LC classes are classified by help of Arc Map 9.3 software.

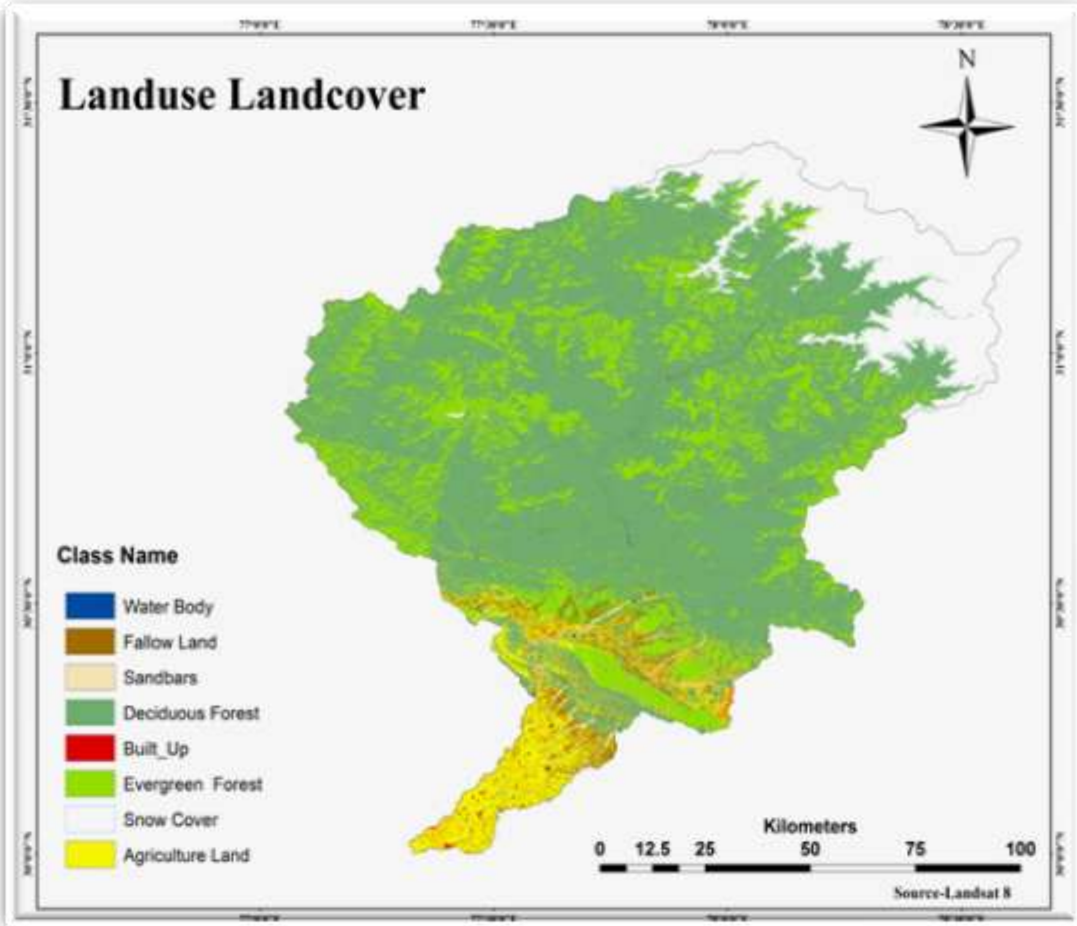


Figure: Land use/ Land cover Map

Rainfall

The earlierresearches shows that flood sizes increase exponentially as a higher fraction of precipitation falls as rain, meaning the size of floods increased at a faster rate than the increase in rain. The impact of rain is directly affect the flood area and its intensity. So the rainfall is a very important element for flood.

The upper basin Yamuna basin and its catchment area is located in Himalayas and monsoon region in north India. India as in monsoon climatic region receive maximum rainfall in the months of May, June, July, August and September. In monsoon season the Yamuna catchment receive high intensity of rainfall cause floods in Yamuna Nagarand its surrounding areas in Yamuna River Basin. These floods affects economy, agriculture, social and cultural life of the people of the Yamuna river basin.

Table: Summary of rain gauge stations in Upper Yamuna Basin

Station	Mean monthly rainfall Sept.2010 (mm)	Minimum monthly rainfall Sept.2010 (mm)	Maximum monthly rainfall Sept.2010 (mm)
Bausan	68.18	1.6	208.6
Haripur	38.00	5.0	233.0
Naugoan	17.82	2.0	94.2
Tuini	13.73	1.6	39.4
Yaswant Nagar	11.5	1.0	65.0
Average	29.85	2.24	128.04

Simulation of Flood with ANN

ANN is a computingsoftware made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs. When the relationships between input data and output data are unknown, they can make a powerful tool for simulation.
The process of ANN flood modelling for forecasting consists of three phases:-
ANN Architecture, training and testing.

ANN architecture

In this work, a three layered ANN architecture consisted of one 1 input layer, one hidden layer and an output layer is used. Seven neurons in the form of elevation, topographic, soil, land use or land cover, geology, flow accumulation, slope and rainfall data, which are the main causative factors for flood in input layer has been used. Only one neuron corresponding to river flow in catchment area in output layer has been used. After input variables and output variables are explained, 7-N-1, ANN architecture is defined as can be seen below:-

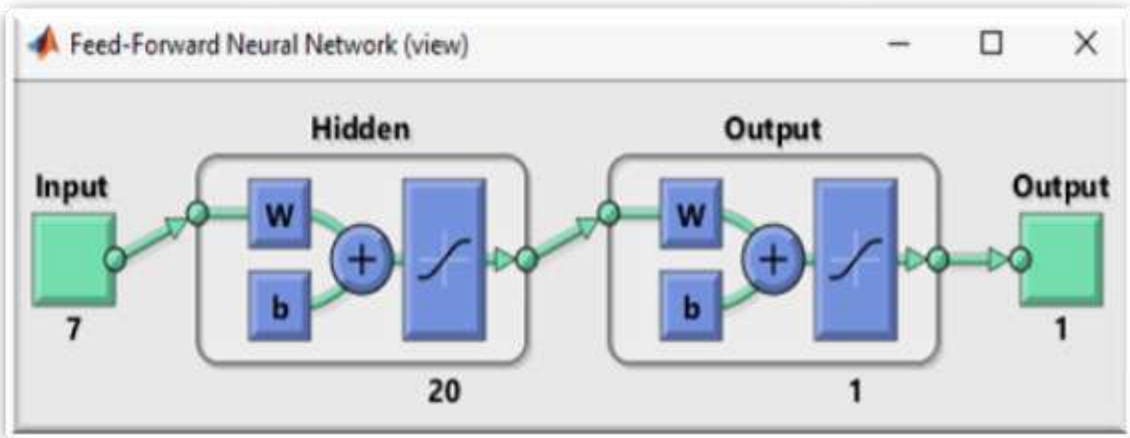


Figure: Feed Forward Neural Network (7-N-1)

Network Training

The goal of train the model or network is to decrease the error betweenthe ANN output and the real data by continuously updating the weightvalues by using Back Propagation algorithm so that model can successfully predict target data from a givenset of input data. Once the minimal error is achieved thetraining process is completed, and the FF algorithm is used toperform required function from whole data set.Here we have applied following training process shown below with flow chart to train our 7-N-1 model which is used in this study, where N represents number of nodes in the hidden layer. Byvarying the number of neurons in both hidden layers, theneural networks are run several times to identify the mostappropriate neural network architecture based on training andtesting accuracies. Most appropriate ANN is selected based onminimum mean square error.

Figure: Training Neural Network (7-N-1)

FINAL INPUT LAYER TO HIDDEN LAYER CONNECTING WEIGHTS

Neurons	Rainfall	Slope	Elevation	Soil	Geology	FlowAccumulation	Land Use
N1	-0.462575142	-1.115516943	0.266820586	-0.411207417	0.108378082	-0.937188135	-0.812018887
N2	1.468772075	0.399388883	0.066288267	0.09119488	0.747769281	0.04088998	-1.926171574
N3	0.626090263	-0.726698764	1.277340015	0.257819074	0.359242936	-0.659136512	1.320008882
N4	-2.397662136	1.371300528	-0.222803945	-0.891362666	0.527176576	0.802062997	-1.887608483
N5	-0.151361024	3.27194541	-0.662481845	0.225097314	-1.837237858	1.334232163	1.142113013
N6	-0.760042696	2.281634224	-0.803984722	1.262843521	0.46381295	0.516043743	-1.509206108
N7	-0.562391083	-0.794982346	-1.334308861	-0.742607768	0.330730244	-0.711900116	0.317796633
N8	1.132496095	-0.78621156	0.78709877	1.659873254	2.121953728	0.532549326	-0.786484669
N9	-2.509853213	0.50921806	-1.971058463	1.255263485	-0.417359382	-1.365574423	1.073858614
N10	0.087962222	-0.776876103	1.858162419	0.470848305	-0.08607241	0.439960062	0.216054885
N11	-0.145924172	-1.500177794	1.727601844	-0.170207341	-1.271565468	0.257321259	0.60751043
N12	-0.470206541	0.372974168	-0.171604546	1.508969999	3.225937047	-1.831060654	1.3901867
N13	1.668548296	1.660050193	1.108047875	1.733537571	-0.402489649	-0.038436849	1.558118401
N14	0.165304815	-0.037842516	0.619328564	1.55772201	-2.097000887	-0.765490782	-1.467838818
N15	-1.387238074	-0.022621891	-0.56998627	-4.743651739	0.692367152	0.532129736	0.658752046
N16	-1.083620142	-0.420743202	0.19292703	1.64752126	0.543381357	0.169890799	1.543844137
N17	0.544548067	-1.985506245	-0.314228738	-0.797780464	-0.424007297	-1.34426439	-0.090261625
N18	-0.308777702	0.437380227	0.090225762	0.742282126	0.03163263	0.216882277	-1.861452804
N19	-0.318151689	0.364957194	2.062633048	-0.840061114	-0.214709461	1.513150662	-1.121802385
N20	-2.185696786	-2.832940265	-0.738043864	-0.094526744	0.902653544	-2.057638037	0.085835741

Analysis results for the input factors

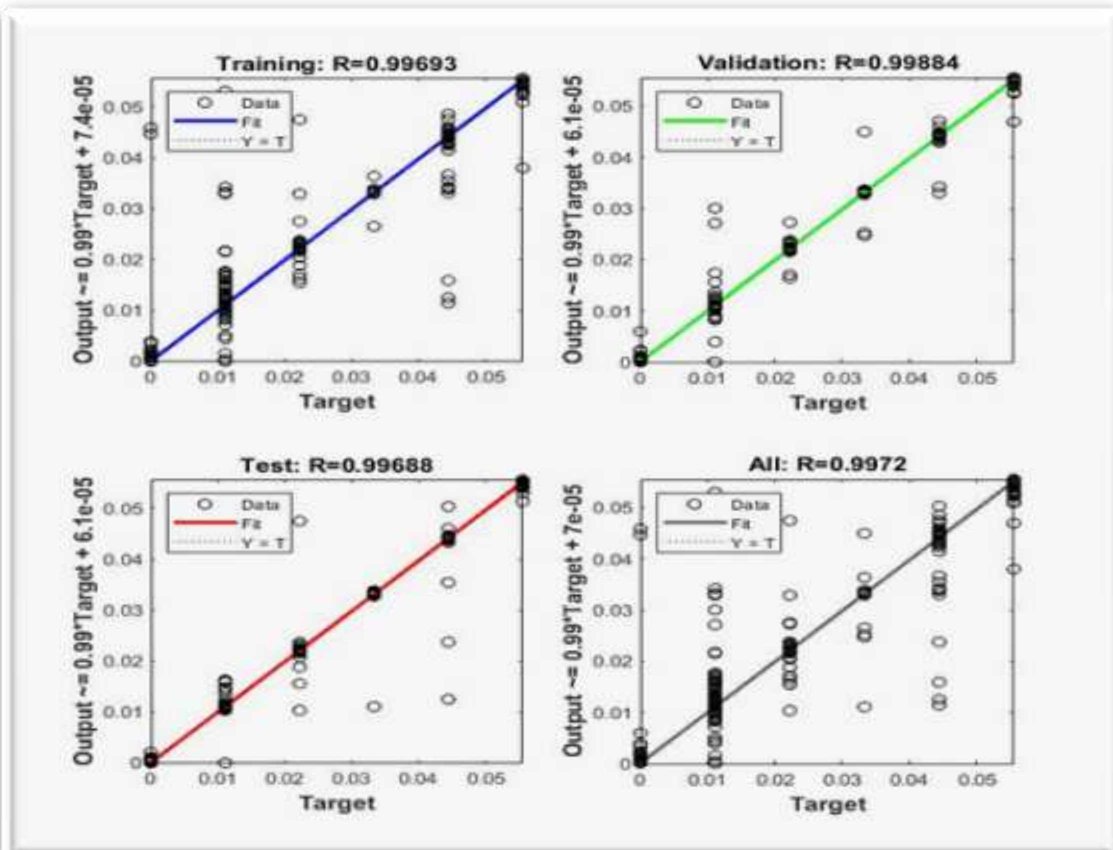


Figure:Diagram of ANN for flood modeling

Network Testing

When training of model is completed, testing data (different datasets) are used to find out the accuracy of model. With the help of new datasets, performance of model was determined. These datasets will have same property as of training data but these are not used for training of the network.

The outcome of testing is measured in terms of R^2 whose value will be equal to 1 or approximate 1,which means result is acceptable i.e. result is showing high level of forecasting. The ANN predicted river flow and regression plot is shown below:

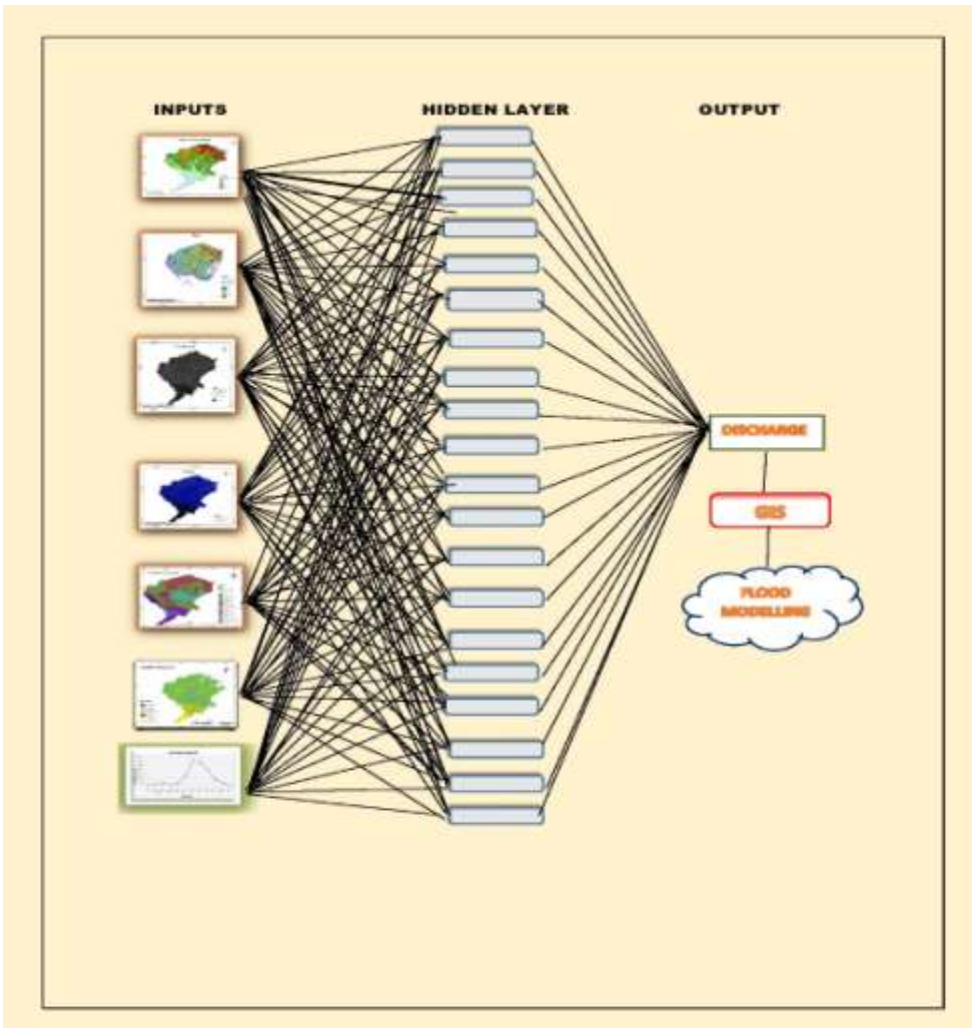


Figure: Regression Plots (R^2)

Flood map setup by applications of ANN model

Maps related to flood help a lot to disaster management for planning in addition to an actual emergency response to the floods. Assessment of flood inundation area is the most important responsibility and most priority for decision makers. The outputs of the ANN model is integrated in GIS for visualization of the flood extent and the flood inundation areas. This map is the best tool to quickly use as a potential impact assessment and for rescue operations for any flood as well as to compute the type and number of buildings affected by the flood. The simulated hydrograph is compared to observed river flow for this event at Hathni

Kund barrage (Fig.5.18). Comparison between the forecasted and observed river flow the accuracy of model is quite good, especially in high river flows. Since the topography is the main factor to specify the flood inundation extent, the DEM map was used to determine the flood-prone area. The flood inundation area is derived from the DEM based on water levels in the river cross-section.

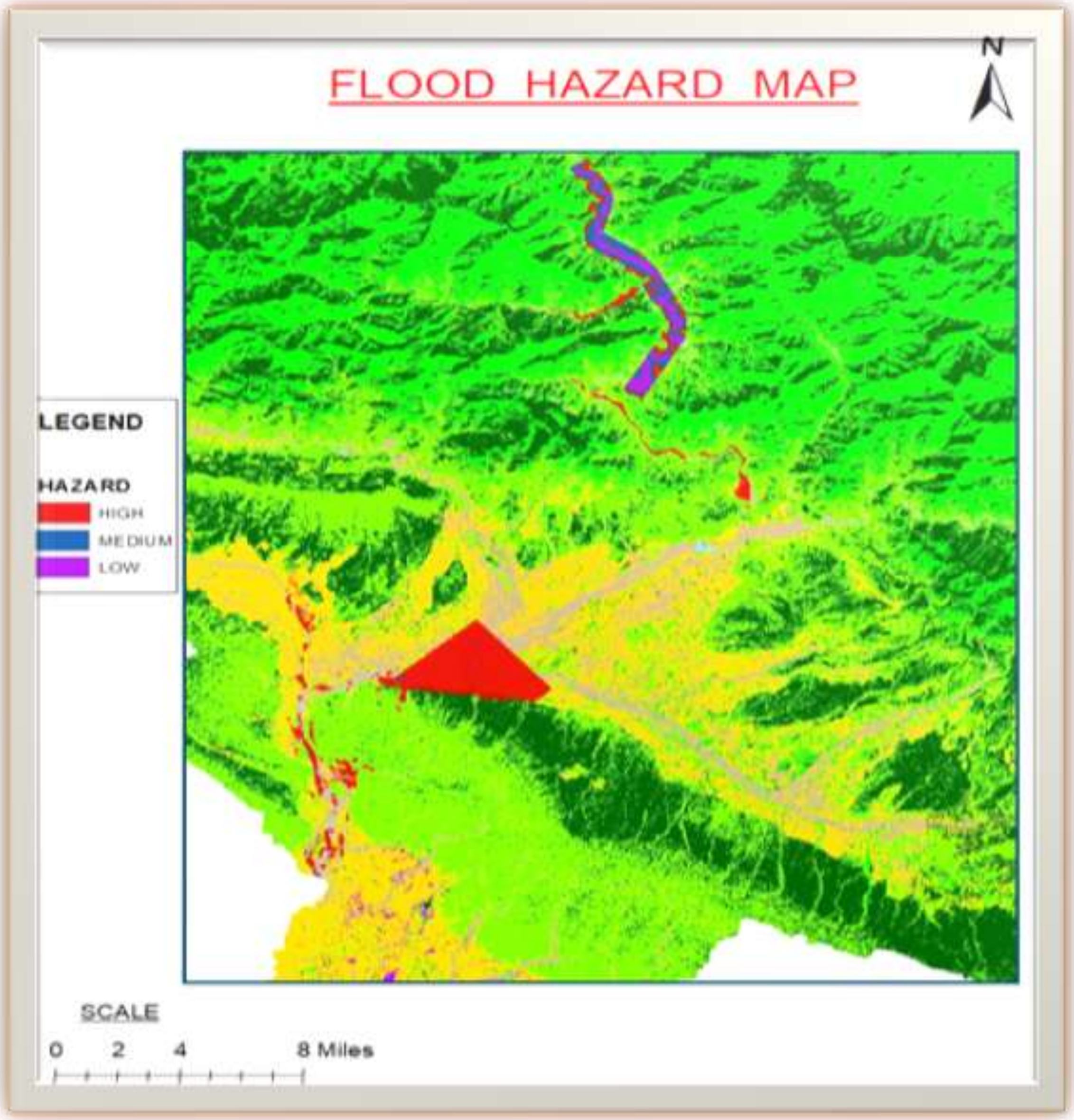


Figure: Flood Simulation Map

III. Conclusion

In this chapter, a flood prediction model to forecast flood in Upper Yamuna River Basin, Haryana using Artificial Neural Network (ANN) has been used. This model is predicting water level not only by taking rainfall as input rather it is considering other causes also because a number of factors can be responsible for changes of water level in river. This model is nonlinear in nature and therefore Multi Linear Perceptron

(MLP) based ANN's Feed Forward algorithm is used. After creation of a database from inputs and outputs; the succeeding phase is the training of the input data through "trainlm" training function that updates weight and bias values according to Levenberg-Marquardt Optimization. Levenberg Marquardt (trainlm) function which is one of the most effective and fast algorithm functions in the toolbox, also it modernizes the weight and bias values with respect to the Levenberg Marquardt and is used by the proposed back propagation network architecture because of its high efficiency.Maximum number of iteration to train the network is set to 1000.The training of the input data is done by the MATLAB software by using the NNTOOL. Training is preceding step to testing of the input data with acceptable error is achieved. The NNTOOL automatically retains 30% of the input data for testing and validation and 70% of input data for the training.

Statistical results showed that data fit well in the model. The analysis revealed that our model is successfully predicting flood water level 48 hours ahead of time.
The model accuracy assessment is described in terms of error of forecasting or the variations between observed and predicted values. Results showed that model has less SSE, MSE and RMSE. Overall errors are negligible. The higher value (close to 1) of R^2 proves that model has excellent agreement with the real data.

REFERENCES

1) Greenlee, D. D. 1987. "Raster and Vector Processing for Scanned Linework." Photogrammetric Engineering and Remote Sensing 53 (10): 1383–1387.
2) Jenson, S. K., and J. O. Domingue. 1988. "Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis." Photogrammetric Engineering and Remote Sensing 54 (11): 1593–1600.
3) Kudrat M.and Nag P.(1998);Digital Remote Sensing,Published by *Ashok kumar Mital Concept Publising Company*, 296pp.
4) Abhijit Paul ,Prodipto Das (2017),International Journal of Computer Applications Technology and Research Volume 3– Issue 7, 473 - 478, 2014, ISSN: 2319–8656.
5) Lillesand T.M. and Kiefer R.W.(1979); Remote Sensing and Image Interpretation, John Wiley and Sons, New York.
6) Siegal B.S. and Gillespie A.R. (1980); Remote Sensing in Geology (New York: Academic Press), PP: 702.
7) Singh B., Dowerah J.(2010); ASTER DEM Based Studies for Geological Investigation around Singhbhum Shear Zone (SSZ) in Jharkhand, India, *Journal of Geographic Information System*,PP:11.
8) Diamond J.T. and Wright J.R.(1988); Design of an integrated spatial information system for multiobjective land-use planning, *Environment and Planning*, PP: 205–214.
9) A.R. Gainguly, "A hybrid approach to improving rainfall forecasts," Computing in Science & Engineering, vol. 4, no. 4, pp. 14-21, 2002.
10) L. C. Chang, P. A. Chen, F. J. Chang, "Reinforced Two- Step-Ahead Weight Adjustment Technique for Online Training of Recurrent Neural Networks," Neural Networks and Learning Systems, IEEE Transactions on , vol. 23, no. 8, pp. 1269-1278, 2012.
11) Atkinson PM, Tatnall ARL, "Neural networks in remote sensing," International Journal of Remote Sensing, vol.18, pp. 699–709, 1997.
12) Bishop CM, "Neural networks and their application,"Review of Scientific Instruments, vol. 65, no. 6, pp.1803–1830, 1994.
13) (2014) R. K. Biswas and A. W. Jayawardena, "Water Level Prediction By Artificial Neural Network In The Surma-Kushiyara River System of Bangladesh".[Online]. Availablehttp://www.icharm.pwri.go.jp/training/master/pubilication/pdf/2009/1.synopsis_mee08177_robin.pdf.
14) Jason Smith ,Robert N. Eli (1995), Neural-Network Models of Rainfall-Runoff Process, Journal of Water Resources Planning and ManagementVolume 121 Issue 6 – November.
15) SONG Lirong, ZHAO Shiwei, LIAO Weilin, WANG Zhaoli (2013),Neural Network Application Based on GIS and Matlab to Evaluation of Flood Risk, International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE 2013),pp.296-299.
16) K. Smith and R. Ward, John Wiley, Chichester (1998), Floods: physical processes and human impacts 1998, earth surface process and landforms book, pp.382.

17) Akyildiz I., W. Su, Y. Sankarasubramaniam, and E. Cayirci (2002), A survey on sensor networks, IEEE Communications Magazine, vol. 40 (8), pp.102–114.
18) Biswajeet Pradhan & Ahmed M. Youssef (2009),Manifestation of remote sensing data and GIS on landslide hazard analysis using spatial-based statistical models,Saudi Society for Geosciences,pp.1-9.
19) Doocy S., Daniels A., & Aspilcueta D. (2009),Mortality and injury following the 2007 Ica earthquake in Peru,American Journal of Disaster Medicine, 4(1),pp.15-22.
20) K. Khosravi, H.R. Pourghasemi, K. Chapi, M. Bahri (2016a),Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models Environ. Model. Assess., 188 (12),pp.656.
21) K. Khosravi, E. Nohani, E. Maroufinia, H.R. Pourghasemi (2016b), A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique,Nat. Hazards, 83 (2),pp.947-987.
22) Norouzi G.,Taslimi M. (2012), The impact of flood damages on production of Iran's Agricultural Sector Middle East,J. Sci. Res., 12 (2012), pp. 921-926.
23) Rajurkar M. P., Kothyari and Chaube U.C. (2002),Artificial Neural Networks For Daily Rainfall—Runoff Modelling,pp.865-877.
24) Kumar Santosh, Das Paul Merina (2014), An approach to low-density polyethylene bio degradation by Bacillus amyloliquefaciens,spirger link,pp.81-86.
25) Patil A.(2012), Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. RNA Biol 9(7),pp.990-1001.
26) Cirstea, M. N., Dinu, A., Khor, J. G. and McCormink, M. (2002), Neural and Fuzzy Logic Control of Daves and Power.
27) Lek S, Guégan J.F. (1999), Artificial neural networks as a tool in ecological modelling, an introduction. Ecol Model 120,pp.65–73.
28) Tien Bui, D., Khosravi K., Shahabi, H.,Daggupati, P., Adamowski, J.F., Melesse, A.M., Pham, B.T., Pourghasemi H.R., Mahmoudi, M., Bahrami, S.,(2019), Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model, Remote Sens. 2019,pp.1589.
29) Papaioannou G., Vasiliades, L., Loukas A., Multi-criteria analysis framework for potential flash flood prone areas mapping,Water Resour. Manag. 2015,pp.399–418. .
30) . Sehgal and C. Chatterjee, "Auto Updating Wavelet Based MLR Models for Monsoonal River Discharge Forecasting," *International Journal of Civil EngineeringResearch*, vol. 5, no. 4, pp. 401-406, 2014.
31) Z.J. Haas, T. Small, "A new networking model for biological applications of ad hoc sensor networks," *Networking, IEEE/ACM Transactions on,* vol. 14, no. 1,pp. 27-40, 2006.
32)] M. B. Kia, S. Pirasteh, B. Pradhan, A. R. Mahmud, W. N. A. Sulaiman, and A. Moradi, "An artificial neural network model for flood simulation using GIS:Johor River Basin, Malaysia," *Environ Earth Sci*. Available:doi:10.1007/s12665-011-1504-z.