

## Zonation of flood prone areas by an integrated framework of a hydrodynamic model and ANN

Ravindra Kumar Singh, Ashish Soni, Satish Kumar, Srinivas Pasupuleti and Vasanta Govind

### ABSTRACT

Limited flood zoning regulations and lack of flood control response units in developing countries make flood problems more severe. This study presents a new framework for categorizing a floodplain into critical risk zones by considering hydraulic and topographical aspects related to flood zoning. The framework was developed by integrating output of the MIKE Hydro River Model with an artificial neural network (ANN) technique which was explored in the lower part of Damodar river basin (Jharkhand, India). A total of nine flood causing factors were selected in three layers of ANN architecture which were optimized by a grid search technique. A confusion matrix was employed to check the unevenness and disproportionality in datasets from which were calculated F1 score values for low (0.815), moderate (0.731), high (0.818) and critical (0.64) zones with best overall accuracy of 75.06%. The results were presented in a GIS environment which shows the model correctly predicted 16, 38, 54 and 24 sites under critical, high risk, moderate risk and low risk zones respectively. Elevation and distance from the river were the most sensitive parameters. Further, this study contributes towards flood susceptibility mapping thereby supporting hydrologists in the course of action and decisions for combating floods in watersheds.

**Key words** | ANN, data driven model, flood zoning, hydrodynamic model, machine learning

### HIGHLIGHTS

- Categorization of flood zones in the lower Damodar river basin.
- Output of 1-D hydrodynamic model is considered for hydraulic parameters.
- Hydraulic parameters along with topographical parameters are taken as input in ANN.
- Carried out accuracy assessment of flood zones.
- Determination of relative importance of each parameter in the flooding zone.

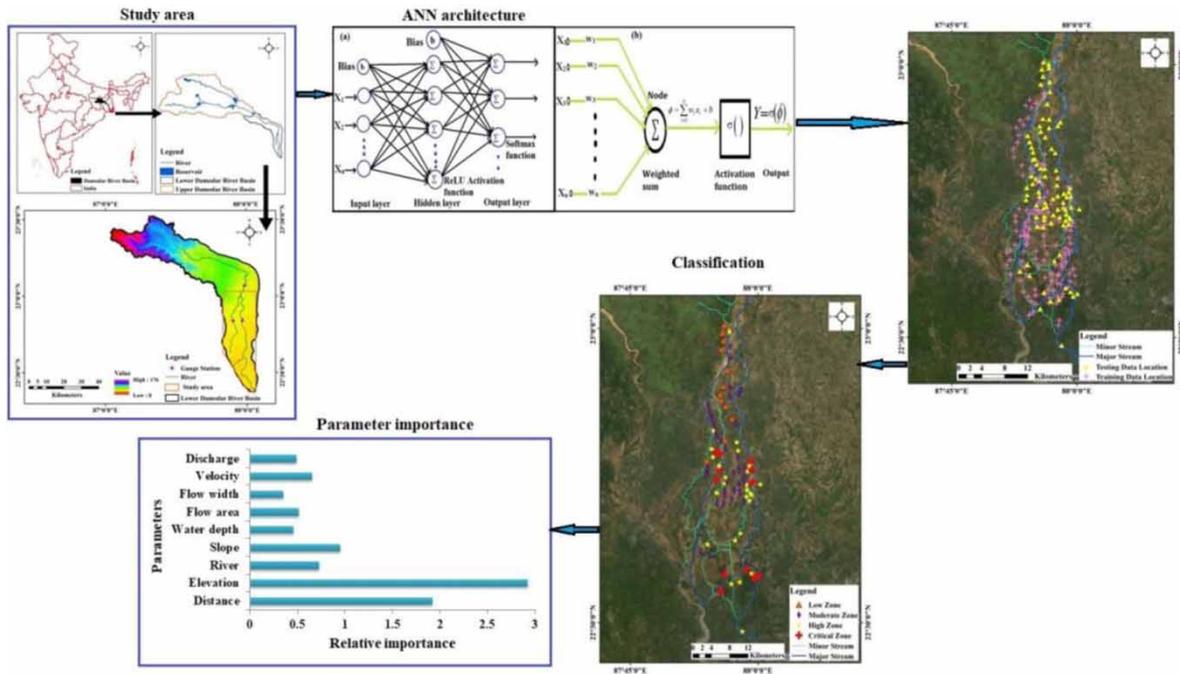
This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

doi: 10.2166/ws.2020.252

**Ravindra Kumar Singh**  
**Srinivas Pasupuleti** (corresponding author)  
Department of Civil Engineering,  
Indian Institute of Technology (Indian School of  
Mines),  
Dhanbad, Jharkhand 826004,  
India  
E-mail: [srinivas@iitism.ac.in](mailto:srinivas@iitism.ac.in)

**Ashish Soni**  
**Satish Kumar**  
**Vasanta Govind**  
Department of Mining Engineering,  
Indian Institute of Technology (Indian School of  
Mines),  
Dhanbad, Jharkhand 826004,  
India

## GRAPHICAL ABSTRACT



## INTRODUCTION

Flood is one of the most common and extensively destructive natural calamities causing huge social and economic damage to various nations globally (Balica *et al.* 2013; Mosavi *et al.* 2018). Coastal, surface and riverine floods have become challenging in recent times due to rapid land-use changes, built-up area growth in flood-prone areas and population growth, along with changing climate (Pranik *et al.* 2010). Absence and lack of proper flood zoning regulations and flood control response units in developing countries makes the flood problems more severe (Solomatine & Price 2004). In developing countries like India and Bangladesh, floods are occurring yearly causing large-scale devastation. So categorizing the zones nearby a floodplain with the uppermost to the lowermost risk and related zone-wise guidelines could be critical for the preparation of upcoming developmental programs, land use policy and give planners an opinion of the regions demanding more or less action throughout a flood disaster (Balica *et al.* 2013; Dwarakish & Ganasri 2015). This flood risk

information is not only important to the policymakers, planners and other authorities but also to the public especially in the affected areas in terms of providing early warnings, evacuation and general preparedness.

Nowadays, for flood risk management, various techniques have been developed to evaluate flood risk and situations investigation for real-time prediction (Balica *et al.* 2013). Physically-based models are extensively used in river hydraulics i.e. 1-D or 2-D hydrodynamic models which are based on mathematical equations that describe the physical phenomena of river flow using spatially distributed parameters. Identification of flooding risk hotspots and hydraulic parameters along a river can be determined with the aid of hydrodynamic models which perform a crucial role in suggesting preventive measures for flood management. The modeling procedure principally comprises the solution of Saint-Venant equations which are partial differential equations which can be used in any practical problem with the help of a numerical method (Chatterjee

*et al.* 2008; Singh *et al.* 2020). Model selection depends upon its functionality and complexity and its performance is controlled by model calibration i.e. identification of their optimum factors which diminish the error between observed and predicted data (Dwarakish & Ganasri 2015).

In the recent two decades, significant growth in computational power machine learning (ML) methods have attracted the attention of modelers and researchers for providing cost-effective solutions in various water resource problems (Sivapragasam *et al.* 2008). Artificial intelligence tools such as genetic algorithms, fuzzy logic and artificial neural networks (ANN) are data-driven (Yamazaki *et al.* 2011). These methods are principally leaning towards optimization (Kim *et al.* 2006). Scholars, through the innovative ML techniques along with hybridization of existing techniques, are moving towards more precise and proficient models (Mosavi *et al.* 2018). ML itself is an innovative data-driven model (DDM) which is the area of artificial intelligence (AI) to have obtained reliable outcomes centred on past statistics and using them to predict the patterns of future behaviour (Mosavi *et al.* 2018). Its incorporation with GIS for water resources has upgraded the capability to generate precise flood models output in a spatial domain. Kia *et al.* (2012) incorporated the ANN technique with GIS for flood modeling the southern part of Peninsular Malaysia. The important ML techniques applied to flood expectancy incorporate ANN, support vector machines (SVM), wavelet neural system (WNN), and multilayer perceptron (MLP) (Nourani *et al.* 2014; Prakash *et al.* 2014; Ashrafi *et al.* 2017; Yu *et al.* 2017). Some researchers used genetic algorithms in optimizing reservoir operation (Wardlaw & Sharif 1999) and predicting temporal scour depth (Pandey *et al.* 2020). Direct incorporation by a professional increases the precision and confidence in the model output (Solomatine & Price 2004; Solomatine & Ostfeld 2008). This technique supports forecasting as it is fast in removing data that is not pertinent and will strengthen the process of evaluating the state of affairs.

ANN is the most prevalent and adaptable among all ML methods and effectively used for several flood forecast applications e.g. streamflow forecasting (Kisi 2007), river flow (Shamseldin 2010), rainfall-runoff (Smith & Eli 1995; Dawson & Wilby 1998) etc. Dawson & Wilby

(1998) had discussed problems that need to be addressed while applying ANN for rainfall runoff modeling. Ghorbani *et al.* (2016) applied ANN and SVM techniques for predicting monthly flow of a natural river (Zarrinehrud river) in north-western Iran. ANN was applied for predicting the peak flow, timing and shape of a runoff hydrograph, based on causal meteorological parameters (Ahmad & Simonovic 2005). Hybridization of ML with other soft computing techniques, numerical simulation or physical models enhanced the performance of model. Hybridization of ML with other soft computing techniques, numerical simulations or physical models enhances the performance of a model. The progression of such novel ML strategies exceptionally relies upon the correct utilization of the soft computing methods in planning unique learning algorithms (Mosavi *et al.* 2018). For precise rainfall-runoff modeling Young *et al.* (2017) incorporated the hydrologic model (HEC-HMS) with machine learning techniques (SVM and ANN). Chang *et al.* (2014) prepared a hybrid ANN model for real-time prediction of provincial flood in a built-up region. The evaluation of the hybrid strategy against each and every single algorithm used in the research demonstrates the usefulness of the proposed technique. It helps planners and managers to improve their understanding and provide various alternatives that support in the policymaking process (Balica *et al.* 2013).

Keeping in mind the usefulness of flood zoning regulations in the preparation of developmental programs and action required during flood disasters, the present study investigated the flood problem of lower Damodar river (West Bengal, India). In the present work, to cater for the flood problem in the lower Damodar river, a new framework was proposed in which output of a hydrodynamic model with other generic topographical parameters was utilized to analyze the flooding zone with the application of ANN techniques. The objective of this paper is to check the performance of the ANN model with the help of a confusion matrix and also to find out the relative importance of flood causative factors. This study contributes to a detailed selection of the most important flood causative factors for flood susceptibility mapping. Outcomes of this study can vastly support engineers and planners in the course of actions and

decisions for combating floods in the lower Damodar basin.

## STUDY AREA AND DATA USED

### Study area

The studies pertaining to flooding have two essential and important components which are selection of study area and data used for the analysis. In the present study, lower Damodar river basin which lies in the state of West Bengal, India was selected for study. The lower Damodar basin is affected by flood during monsoon season due to huge discharge coming into the Damodar river from upper catchment (due to high runoff generation) and excess water release from a dam. Damodar river splits into two important channels, the Mundeswari and the Amta Damodar in the lower segment of the basin. Topography in these areas is flat alluvial plains and also the river carrying capacity is low due to siltation over time. The area from the bifurcation point of the Damodar river to a downstream point was chosen for the present study purpose. A large portion of discharge flows through the Mundeswari river which runs down into the Rupnarayan river. Mundeswari and the Amta Damodar can't convey the total discharge of flood of the Damodar and subsequently, the elbow zone of the river gets immersed. For the present study, the area from the bifurcation point of the Damodar river to the last downstream point was selected because this whole area is a flood prone area. A study area map with geographical extent is shown in [Figure 1](#).

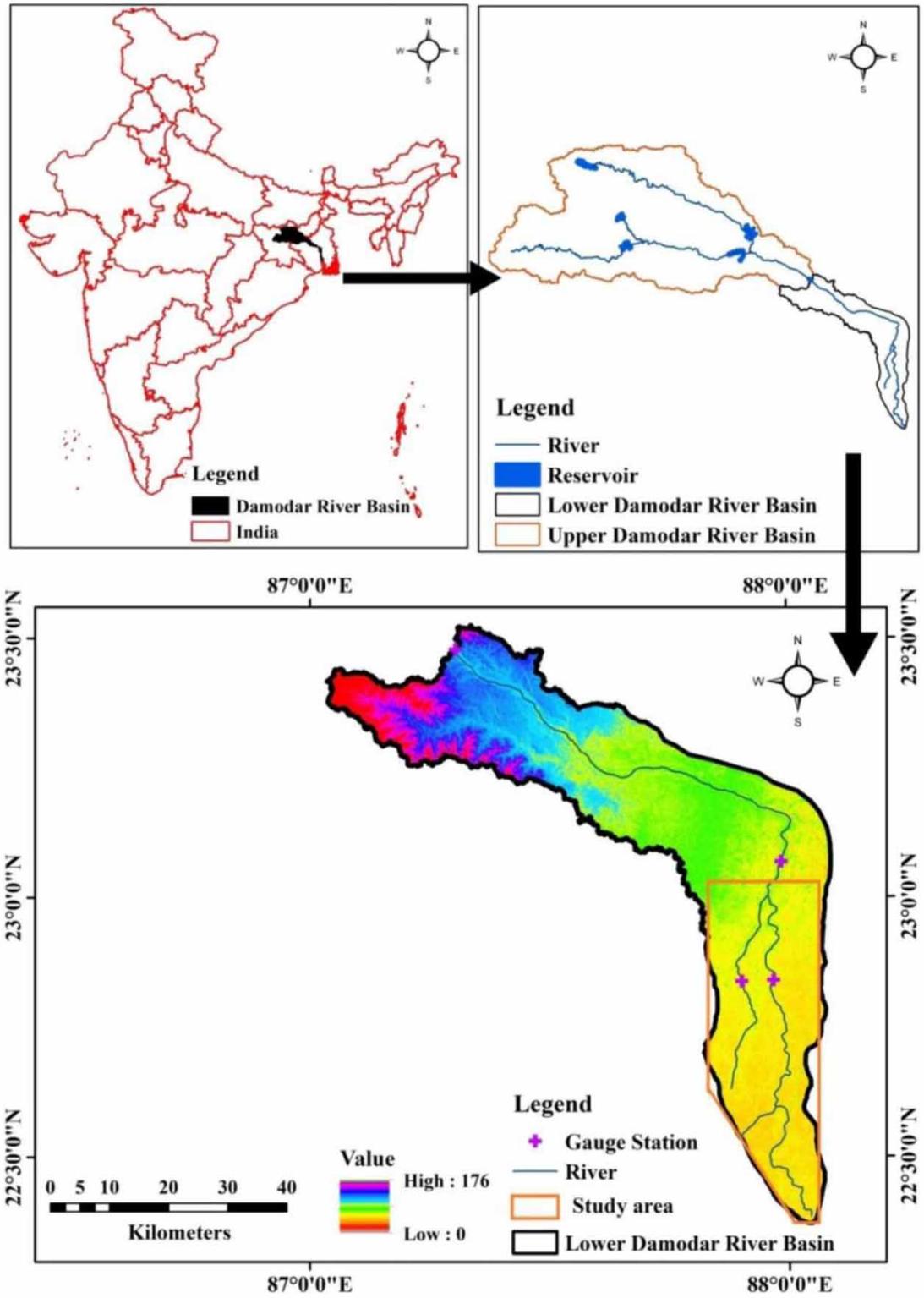
### Data used

For the flood study, two types of data are important namely hydraulic data and topographic data. These data have substantial effects on flooding events. Hydrodynamic model peak annual discharge data and daily discharge data for the Durgapur Barrage were collected from the Hydraulic data division Damodar Valley Corporation (DVC) Maithon, Jharkhand. Gauge data measured at the Jamalpur, Harinkhola and Champadanga gauging stations were collected from the Central Water Commission (CWC), and

the Irrigation and Waterways Department of West Bengal. Cartosat-1 with spatial resolution of 2.5 m, LISS IV with spatial resolution of 5 m and Cartosat-1 Digital Elevation Model (DEM) with spatial resolution of 10 m were procured from the National Remote Sensing Centre Hyderabad for geometry data preparation in the model. Landsat satellite imagery with spatial resolution of 30 m was used for identification of past water logged areas during flood time and Google earth images were used for better visualization in the study. Cartosat DEM was modified based on error analysis of the DEM before using in the hydrodynamic model. The error analysis of the DEM was performed by comparing the elevations of the DEM and spot heights along with field data. From the error analysis, it was decided to adopt a simple method of modifying the DEM extracted cross-sections by subtracting the values of  $RMSE_{DEM}$  from the elevation values of the extracted cross-sections. Modification of the cross-sections by subtracting the RMSE values might not ensure complete matching of the geometry but the elevations of the river banks became closer to their corresponding observed values. For the ANN method hydraulic output data from the developed 1-D hydrodynamic model of lower Damodar river ([Singh \*et al.\* 2020](#)) was used. Hydraulic data consists of various parameters namely discharge, velocity, water depth, flow width and flow area at different cross-section locations in the river. Topographical data derived from DEM were used to understand topographical influence on the spatial extent of flooding. Distance from river bank, slope pattern, elevation and river class were used as topographical data.

### Brief description of 1-D hydrodynamic model

In the present study, for the ANN technique hydraulic parameters along different cross-section locations in the river for flood zoning were taken from the output of a 1-D hydrodynamic model (MIKE HYDRO RIVER). [Singh \*et al.\* \(2020\)](#) performed the 1-D hydrodynamic model for flooding time in lower Damodar using the MIKE HYDRO RIVER model for the time period 1st July to 17th October 2007 and from 1st July to 15th October 2009. The model solved the real field problem of hydraulic domain by solving Saint-Venant equations with the help of implicit finite difference



**Figure 1** | Geographical location of lower Damodar river highlighting study area and river.

numerical schemes. Saint-Venant Equations are a combination of conservation of mass and conservation of momentum equations and are presented in Equations (1) and (2) for 1-D flow.

Continuity equation

$$\frac{1}{T} \frac{\partial Q}{\partial x} + \frac{\partial y}{\partial t} = 0 \quad (1)$$

Momentum equation

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left( \frac{Q^2}{A} \right) + gA \frac{\partial y}{\partial x} + gA(S_o - S_f) = 0 \quad (2)$$

where  $Q$  = discharge ( $\text{m}^3/\text{s}$ ),  $T$  = top width (m),  $y$  = water depth (m),  $x$  = longitudinal distance (m),  $g$  = acceleration due to gravity,  $t$  = time step (sec or h),  $S_f$  = energy slope,  $S_o$  = bottom slope,  $A$  = flow area (cross-sectional area).

The MIKE HYDRO RIVER hydrodynamic model used six-point Abbott and Lonescu implicit finite difference numerical schemes to solve the Saint-Venant equations. The computational efficiency of numerical simulation is vastly reliant on the computational grid of the model applied during simulation. During analysis the model automatically generated alternate discharge points and water level points in a computational grid. 476 cross-sections were extracted from high resolution DEM (Cartosat-1) by the model for performing the simulation. Taking care of the river's physical variations in shape and mathematical demands 158 cross-sections on the lower Damodar river from Durgapur barrage to the bifurcation point of Damodar river, 183 on Amta Damodar, 117 on Mundeswari river and 18 cross-sections on the branch of Amta Damodar river were selected. Model calibration was performed to match the observed and simulated gauge data by variations in the value of Manning's roughness coefficient ( $n$ ) which was in the range of 0.02–0.045. The calibrated result was best performed for Manning's  $n$  value of 0.03. All the data related to hydrodynamic model simulation is presented in Table 2.

For more details please refer to Singh *et al.* (2020). Output from this 1-D hydrodynamic modeling study of lower Damodar of Singh *et al.* (2020) in the form of hydraulic data was used in the present ANN model along with

other topographical data. Various data used in the ANN technique with range are shown in Table 2.

## RESEARCH METHODOLOGY

This section discusses the methodology used to detect flood prone zones in the lower Damodar basin. Figure 2 illustrates the sequential architecture to the proposed methodology. The architecture consists of three main parts: (a) Data preparation (b) Data pre-processing (c) Modeling process. The selection of data and various factors are discussed in the data preparation section. To fit the data and understand the statistics of the different parameters, transformation has been performed in the 'data pre-processing' section. Lastly, the modeling process section defines the ANN technique and its optimization for efficient prediction.

### Data preparation

In the present study, hydraulic parameters namely discharge, velocity, water depth, flow width and flow area at different cross-section locations in the river were used to interpret their effects on the flooding zone. Topographical data derived from DEM to understand topographical influence on the spatial extent of flooding consist of distance from river, slope pattern, elevation of area and river class. The distance of an area from the river has a significant effect on flooding. Chance of flooding is higher in an area which is near to the river compared to an area which is far away. Slope pattern of any area is divided into three parts i.e. slope pattern near to riverbank, slope pattern at midpoint from riverbank to the particular area of interest and slope pattern near to the particular area. Falling slope is denoted by 'F' whereas rising slope is denoted by 'R'. For an area (under flooding) having a falling slope pattern from the river, there is a higher chance of flooding and if a slope pattern is rising then there is less chance of flooding. If the slope pattern between the river to any area is a combination of both types then other factors played a significant role in deciding the zone. Hence slope was categorized into four types i.e. F, FFR, FRF and RFF means falling, rising near area but falling at other points, rising near at midpoint but falling at rest of the part and last rising near river

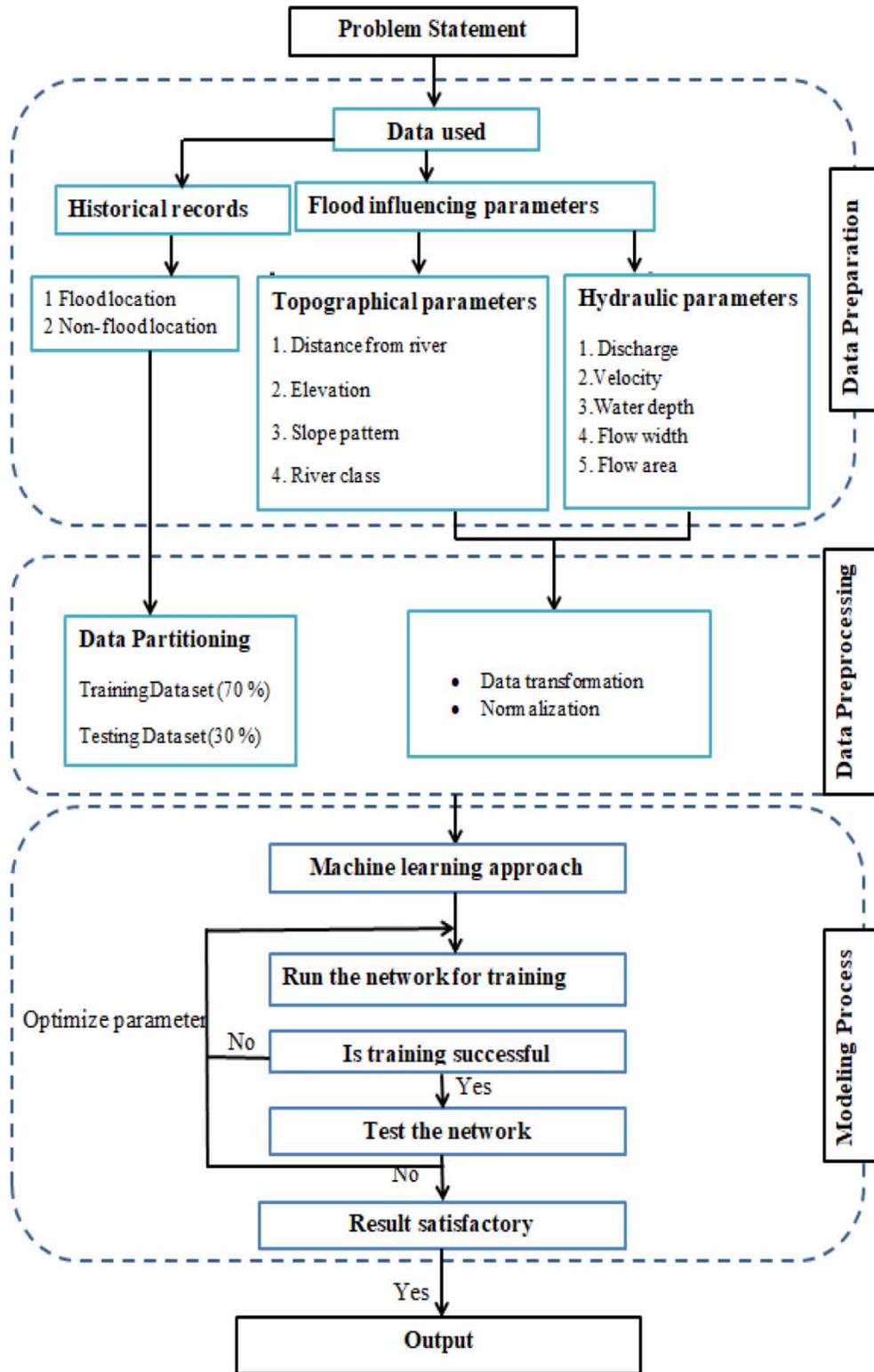


Figure 2 | Flowchart of the applied methodology.

**Table 1** | Model simulation criteria

Simulation criteria	Values
Simulation period	1st July to 17th October 2007 and 1st July to 15th October 2009
Cross-section spacing	205 m–1,270 m
Total cross-sections	476
Time step	1 minute
Output data storage frequency	180 time steps = 180 minute
Upstream Boundary condition	Daily Discharge Data
Downstream Boundary condition	Q/H curve
Initial condition	0.5 m water depth
Manning's roughness coefficient (n)	0.02–0.045

bank but falling elsewhere. Elevation of area means whether the particular area is at a higher or lower level. Possibility of flooding is high in lower elevation compared to a higher elevation. River class means which river initiates the flooding in a particular area. The Damodar river is bifurcated into the Mundeswari and Amta Damodar and also one branch of the Amta Damodar is created downstream of the Amta site for taking care of extra discharge of the Amta Damodar river. So to know the impact of a river, a different numeric value is assigned to each one: class 1 for the Mundeswari river, class 2 for Amta Damodar and class 3 for Amta Damodar Branch. Also, the investigation of historical flood instances and location records are

necessary for evaluating flood hazard in a region. The flood inventory map in this study was principally formed by representing the flood sites where floodwater overtopped the river and inundated the area by utilizing the ground survey and old government records.

### Data pre-processing

Pre-processing of the dataset is an important task to implement the machine learning model. Handling the missing values and transformation was performed in this proposed methodology under data pre-processing. Handling missing values was performed in the dataset where a data point was unavailable due to various reasons, typically replacing it with the mean value.

Transformation is the process of changing the values in the dataset for the compatibility of machine learning algorithms. In this study categorical and feature scaling were performed for transformation. Categorical transformation converts non-numeric data points (like slope pattern as in Table 1) to numeric data points. This transformation changes the string dataset to machine learning format, which make it easy for algorithms to perform the matrix multiplication operation.

The most important process in any machine learning algorithms, is the feature which generally normalizes the range of raw data points into 0–1. For example, water depth, flow width, flow area etc., have a different range of values and the sharing weight of the ANN neuron have values in the range

**Table 2** | List of data involved in ANN model with range

S.N.	Parameter	Training data range	Testing data range
1	Water depth (m)	2.36–9.46	2.36–9.87
2	Flow area (m <sup>2</sup> )	171.28–6610.67	196.8–6292.33
3	Flow width (m)	186.93–2682.85	196.41–2418.54
4	Velocity (m/s)	0.031–6.67	0.05–4.58
5	Discharge (m <sup>3</sup> /s)	37–10187.21	63.64–9788.63
7	Distance from river (m)	150–4500	50–4290
6	Elevation	3–17	4–21
8	Slope pattern	F, FFR, FRF, RFF	F, FFR, FRF, RFF
9	River class	Mundeswari (1), Amta Damodar (2), Amta Damodar branch (3)	Mundeswari (1), Amta Damodar (2), Amta Damodar branch (3)

of 0–1. This situation causes abnormal gradient change while training the model, and very high variation in the multiplication factor. Therefore, the range of all attributes of the dataset should be normalized so that each feature contributes approximately proportionately. Here, Min-Max normalization has been used to make the feature proportional. The general formula for normalization is shown in Equation (3):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where  $x$  is the original value,  $x'$  is the normalized value.  $x_{min}$  and  $x_{max}$  are the minimum and maximum value of a particular attribute in the dataset.

### Modeling process

In the present study, an ANN technique was used for classifying the flood zones of lower Damodar river. The ANN architecture refers to a highly compact design of interconnected neurons to transfer information and share weight among them. Designing a suitable architecture is a tedious and also important task. The theoretical description of ANN was described by an investigator (Haykin 1999). Use of an ANN model is comprised of three main steps: first one designing the ANN architecture, second one is optimizing the network parameter while training and third one testing the unknown dataset for validation of the model.

Each layer consists of numerous nodes or neurons (Jahangir et al. 2019). The ANN architecture denotes the number of layers involved and connection weights. Various researchers suggest a different architecture for different input; however, there is no rule to define the architecture of an ANN. A basic overview of ANN architecture is shown in Figure 3(a). Each neuron of an ANN model receives the information from the number of input parameters and these inputs multiply by their corresponding weights. A neuron adds the values of all inputs as shown in Figure 3(b). An additional term 'b' is included during summation which is known as bias. An activation function is applied to the summation values to generate the final output. The activation function has a major effect on the neural network's convergence speed. It is attached to each neuron in the network and decides whether each neuron's input is significant for the model's prediction or not. The generalized formula for a neural network having multiple layers and multiple nodes in each layer is shown in Equation (4). Here, the ANN comprises of an input, two hidden and one output layer. The input layer was designed to take the input of nine factors in the present study that are responsible for flood (shown in Table 2). A Rectified Linear Unit (ReLU) activation function was used in the present study and shown in Equation (5). The output layer contains a single neuron to calculate the probability of each class via a softmax function then chooses the higher probable class as output which is shown in Equation (6). This function is used when there is

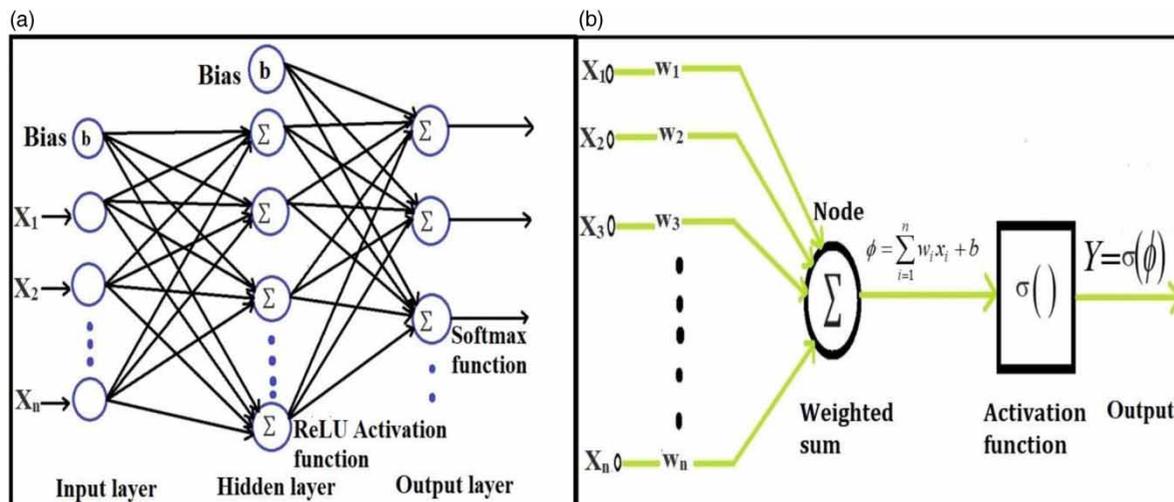


Figure 3 | (a) Pictorial representation of the ANN architecture (b) Pictorial representation of the individual node processing.

multiclass classification and it turns the number into probabilities which sum to one.

The generalized formula for an ANN having multiple hidden layers and multiple nodes in each of the layers i.e.  $L$  layers with  $n$  nodes and  $L-1$  layer with  $m$  nodes and so on, is:

$$a_n^L = \left[ \sigma \left( \sum_m w_{nm}^L \left[ \dots \left[ \sigma \left( \sum_j w_{kj}^2 \left[ \sigma \left( \sum_i w_{ji}^1 x_i + b_j^1 \right) \right] + b_k^2 \right) \right] \dots \right] + b_n^L \right) \right]_n \tag{4}$$

where  $w$  = weight;  $b$  = bias;  $x$  = input;  $L$  = layer;  $i, j, k, m, n$  = node in corresponding hidden layer,  $\sigma$  = activation function;  $\sum$  = weighted sum.

ReLU activation function ( $\sigma$ ) is:

$$\sigma_{(x)} = \max (0, x) \tag{5}$$

It gives an output  $x$  if  $x$  is positive and 0 otherwise

The Softmax Function is:

$$S(z)_j = \frac{e^{z_j}}{\sum_{k=1}^k e^{z_k}} \quad \forall \quad j = 1, 2, \dots, k \tag{6}$$

where  $k$  is the dimension of vector  $z$  of arbitrary real value.

In this study, the grid search technique was used to get efficient ANN architecture. The grid search technique was used to handle the tuning parameter of the machine learning model during the training process which is also known as hyper-parameter optimization. Herein, different tuning parameters (also known as hyper-parameters) were used like learning rate, activation function, number of neurons in each layer and batch size. The main objective of the machine learning approach is to reduce the gap between predicted and actual output by changing the weight of the neuron in the hidden layers during the training stage. The gap is also known as error in machine learning terminology. The changing weight is typically done by back-propagation algorithms. In this study, a stochastic gradient descent (SGD) algorithm was used which reduces the sharing weight of the neuron by calculating the error derivative with respect to each neuron. The loss/error function was used to evaluate the accuracy of model during the training stage. To get the optimal performance and real-time assessment of the model, early stopping was used in this study. These approaches stop the training process as soon as the minimum error is achieved or error curve converges. In

the present the study maximum number of iteration was terminated at 200 epochs. To achieve the unbiased performance of the model with respect to the dataset, the data were randomly selected with a ratio of 70:30, 70% dataset was used to train the model. The remaining 30% is future split into 20:10, where 10% is used for the validation of the model to avoid the over-fitting problem, whereas 20% is used for testing the model. The data location along with major and minor stream is shown in Figure 4.

Interpretation from an accuracy metric alone can be misrepresentative if each class has disproportionate data or targets have multiple classes in the dataset. So in the present study a confusion matrix was used to verify the ANN’s performance in terms of classifications. Along with accuracy metric, other evaluation metrics like precision, recall (sensitivity), specificity and F1 score were calculated from the confusion matrix. Specificity is the metric that evaluates a model’s ability to predict true negatives (TN) of each available category. Selection of recall is better if the idea of false positives (FP) is far better than false negative (FN). Precision is very useful if the modeler wants to be more confident about true positives (TP) whereas specificity is a better choice if the modeler wants to cover all true negatives (TN). F1 is typically more valuable than accuracy in uneven class distribution because it is the weighted average of precision and recall. The equations for calculating these metrics are shown in (Equations (7)–(11)):

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{7}$$

$$\text{Sensitivity (Recall)} = \text{TP}/(\text{TP} + \text{FN}) \tag{8}$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \tag{9}$$

$$\text{Accuracy} = \text{TP} + \text{TN}/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{10}$$

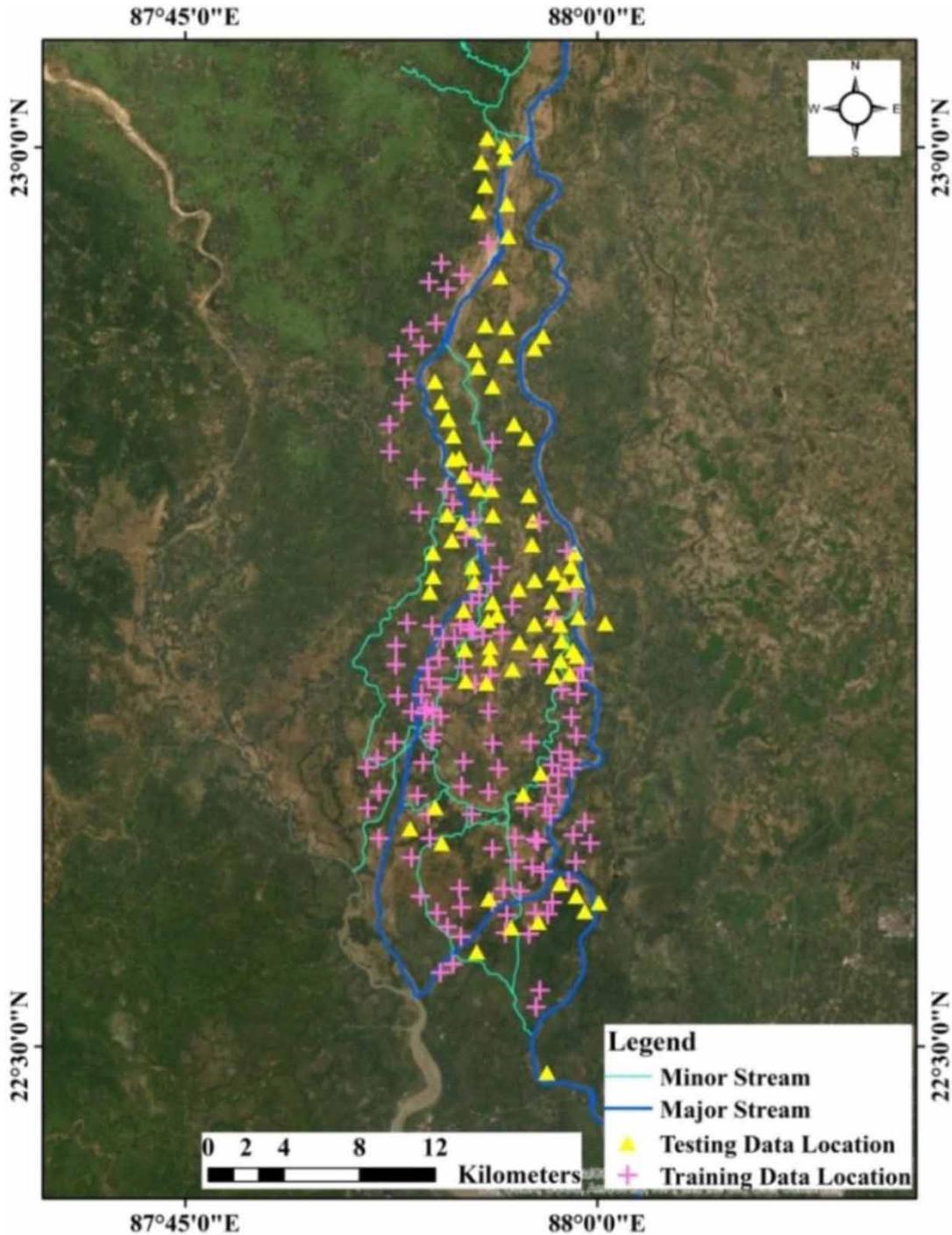
$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision})/(\text{Recall} + \text{Precision}) \tag{11}$$

TP = True positive means model correctly predicted the positive class.

TN = True negative means model correctly predicted the negative class.

FP = False positive means model incorrectly predicted the positive class.

FN = False negative means model incorrectly predicts the negative class.



**Figure 4** | Data locations along with major and minor streams.

The whole experiment was designed using Python environment, where TensorFlow library was utilized as the backend framework to run the machine learning algorithms.

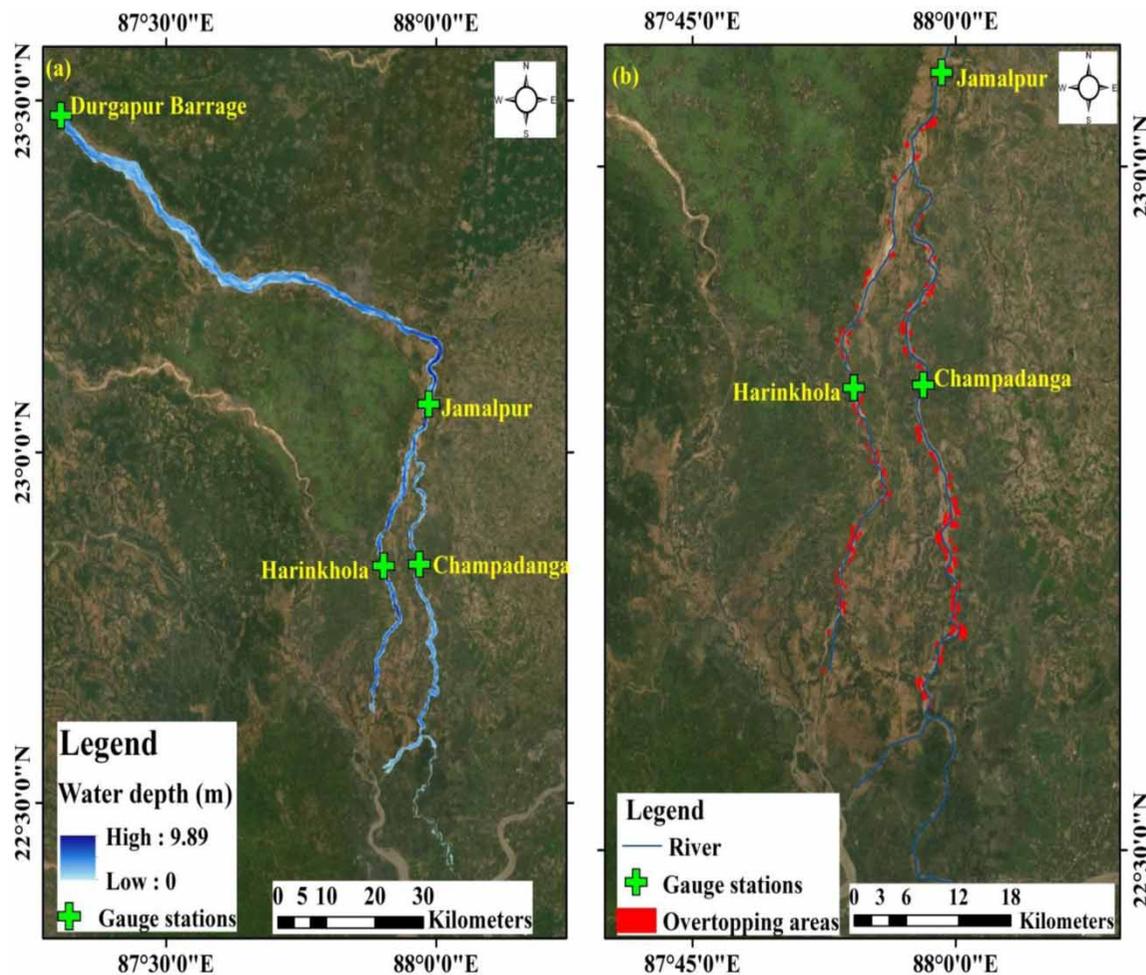
Keras as well as sklearn were also involved in the modeling for different pre-processing as well as post processing tasks. Matplotlib library was used to generate modeling graphs.

## RESULTS AND DISCUSSION

### 1-D hydrodynamic model

The hydrodynamic model produced satisfactory output at the gauging station of lower Damodar river for Manning's roughness ( $n$ ) value of 0.03 which was very close to observed field data. A water depth map was prepared using the output of the model and with the help of this map overtopping areas nearby the river were demarcated as shown in Figure 5(a) and 5(b). Overtopping areas which were demarcated based on the model output represent only those areas which area under the model domain. Areas downstream to the branching point of the Damodar river are more at risk of flooding as identified by the

model output and as satisfied by field validation. Mundeswari and Amta Damodar river cannot accommodate the excess flow of upstream catchment throughout the flooding time. As a result, excess flow will overtop both banks of the river and consequently, the area adjoining the river gets affected as shown by red spots in Figure 5(b). Due to presence of large plain areas water effortlessly spreads in the major part of the region after overtopping the bank of the river. For more details please refer to Singh *et al.* (2020). So, considering the above issues advanced study using ANN for the area between Mundeswari and Amta Damodar river as an extension of the hydrodynamic model (Singh *et al.* 2020) work was performed in the present study. Output from the earlier hydrodynamic modeling study of lower Damodar in the form of hydraulic data parameters



**Figure 5** | (a) Maximum predicted water depth map overlaid on Satellite imagery (b) Demarcated overtopping areas based on model overlaid on satellite imagery (Source: Singh *et al.* 2020).

which shows the response of the river at different cross-sections to the excess flow was used in the present ANN model along with other topographical data. Hydraulic parameters of only the catchment area surrounded by Mundeswari and Amta Damodar river were taken from the hydrodynamic model for analysis in the present ANN model.

## ANN

The present ANN model performs the iterations for classifying the flood zone in different classes based on the earlier mentioned hydraulic data and topographic data (derived from high resolution Cartosat-1 DEM, 10 m × 10 m). The target of the dataset consists of various flood zones i.e. low risk, moderate risk, high risk and critical risk across the region. A constant value of learning rate (LR) i.e. 0.0001 was used during the training process and also to accelerate the algorithm iteration, a momentum constant (Nesterov's momentum) of 0.9 value was used in the training process. SGD increases the iteration process faster because it uses only one or a subset of a training sample from the training dataset and starts improving itself right away from the first sample. This increases the performance and reduces the training time of the model. In the present study the maximum number of iterations was terminated at 200 epochs. To get an efficient prediction by the model, an iteration vs loss graph was used. It was observed from Figure 6 that the error decreases with iteration. The model predicted the flood zone into critical, high risk, moderate risk and low risk flood zones with an overall accuracy of 75.6%.

In the study, a confusion matrix was used to measure the performance of the ANN model in terms of classifications. The numbers of correct and incorrect predictions is summarized by each class in the confusion matrix and is shown in Table 3. Elements of each box of the confusion matrix in rows and columns have their own significance. Diagonal elements of the confusion matrix show the true positive value of each class (flood zone) means the model correctly predicted the target value. Columns of the confusion matrix represent predicted values of the model whereas rows represent actual values. For example 16 and 8 values in the column of critical zone shows the model correctly predicted 16 sites of critical zone (true positive) and 8 sites predicted as critical zones but those sites actually belong to the high zone (false positive).

In the present study datasets were disproportionate in each class due to which the accuracy metric did not give a clear idea of model performance. So other metrics, namely precision, recall, specificity and F1 score were calculated for a clear evaluation of the model performance. Zone wise, the value of each class i.e. true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are separately presented in Table 4. Performance metrics for all classes were calculated based on those values. The precision metric is used when the confidence of a false positive is high whereas the recall metric is used when the confidence of a false negative is very significant. But for the flooding zone classification, there is a need of balance between both the metrics (precision and recall). If the model incorrectly predicted critical flood zones being in a low zone then people could settle in these areas and perform

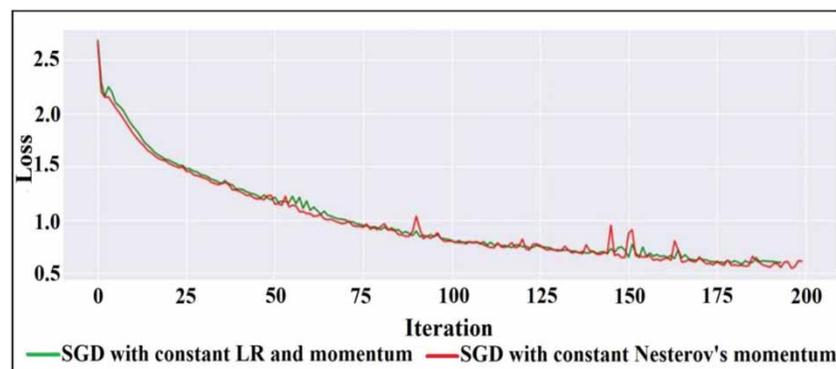


Figure 6 | Iteration vs loss graph.

**Table 3** | Confusion matrix

	Low Zone	Moderate Zone	High Zone	Critical Zone
Low Zone	22	6	0	0
Moderate Zone	4	54	4	0
High Zone	0	10	38	8
Critical Zone	0	0	10	16

**Table 4** | Performance metrics of model

	Low Zone	Moderate Zone	High Zone	Critical Zone
TP	22	54	38	16
TN	108	76	92	114
FP	4	16	14	8
FN	6	8	18	10
Precision	0.846	0.771	0.731	0.667
Recall	0.786	0.871	0.678	0.615
Specificity	0.964	0.826	0.868	0.934
Accuracy	0.928	0.844	0.802	0.878
F1 Score	0.815	0.818	0.731	0.64

some development work. If flooding then occurred in those areas huge loss of life and economic loss could result because in actual fact these areas come into the critical zone. Also if the model incorrectly predicted low flooding zone areas to be in a high or critical zone then in that case also society and government suffer losses because there is no proper utilization of land. So F1 score might be a better metric to use for model evaluation in the present study as it is the weighted average of precision and recall. From [Table 4](#) it was observed that in the case of critical zone, difference between the values of precision and recall was not very high. This means the difference between incorrect predictions of some sites as critical zones (8 sites) which were actually some other zone (high zone) and also prediction of some critical zone sites (10 sites) into different a zone (high zone) by the model was not very high. This balance was also reflected in F1 score (0.64) values of the critical zone. But the best part of the model prediction was that it incorrectly predicted sites only in their adjacent zone (in terms of danger) not in a zone which was very different, such as a critical and low zone pair. So severity of flood zone was still maintained high as predicted by the model.

Same things were observed for all flood zones. For critical zone precision, recall was on the lower side whereas specificity was higher, which shows a high true negative means correct prediction of another class. For moderate zone recall was higher than specificity whereas for other classes the trend is opposite. This means for the moderate zone percentage of model prediction for true value was higher (4.5% higher) than true negative whereas for other classes prediction for true negative was higher (17.8% for low risk zone, 12.3% for high risk zone and 31.9% for critical zone). Predicted classified flood zones calculated by the ANN model overlaid on satellite imagery is shown in [Figure 7](#) which depicts that all the flood zones were closed to major and minor streams which were overtopped during flooding times and inundated all the areas.

Critical zones were identified by the model along all three streams namely Mundeswari, Amta Damodar and Amta Damodar branch. The parameters for the critical zone vary from stream to stream. In the case of Amta Damodar the predicted critical zone maximum distance of the site from the river was 450 m and elevation for this point was 7 m. On the other side, if maximum elevation was considered it was 13 m and for this elevation distance of site from the river was 200 m. For the Mundeswari river the maximum distance for the predicted critical zone was 390 m and at this point elevation of the site was 9 m. Whereas maximum elevation for the predicted critical zone in Mundeswari river was 12 m and distance from the river was 250 m. In the case of Amta Damodar branch for critical zone, maximum distance was 760 m and elevation for this point was 5 m. For larger distance water depth has higher values and flow area has lower values for greater water depth. These hydraulic values represent values at the nearest river cross-section of a critical site. So a higher elevation very close to a river comes under a critical zone and also a very low elevation site at some larger distance if suitable downslope is present. Hydraulic values were on the higher side in Mundeswari river compared to Amta Damodar. So knowledge of parameter importance in deciding the flooding zone is very crucial.

Some researchers have investigated the relative importance of each parameter causing the flood to improve the model ([Kia \*et al.\* 2012](#); [Tehrany \*et al.\* 2015](#); [Rahmati & Pourghasemi 2017](#)). Sensitivity analysis is a general technique for searching essential model parameters, checking the model

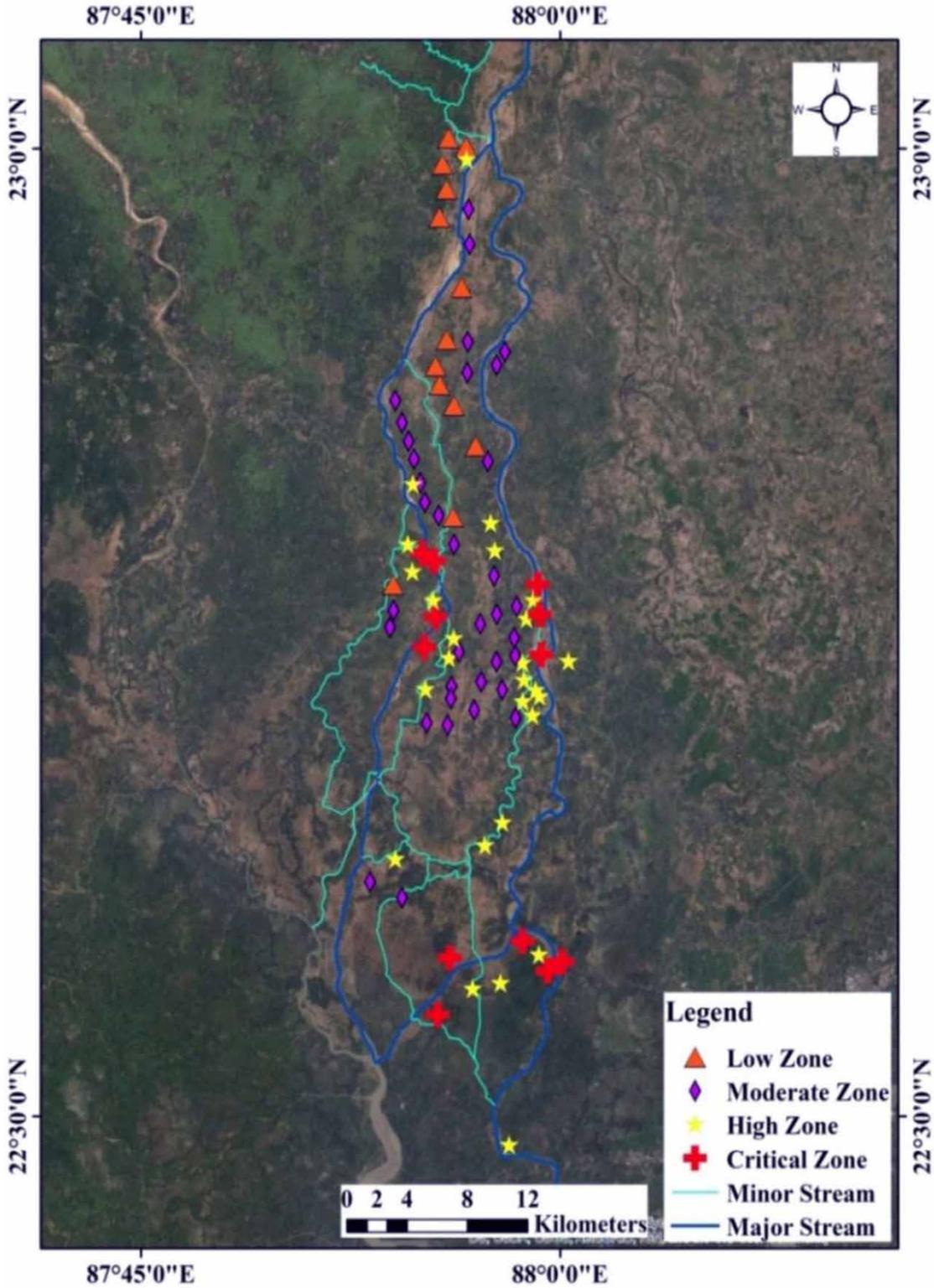


Figure 7 | Predicted flood zone by ANN overlaid on satellite imagery.

conceptualization and evolving the model structure. In this study influence of each parameter was calculated and shown in Figure 8. From the influenced plot it was found that elevation of the surface and distance from the river influenced flooding to a greater extent. Afterward, slope has higher influence on flood zoning compared to other factors as can be seen in Figure 8. Influence of all the hydraulic parameters compared to topographical parameters as predicted by the present ANN model was less. The reason for this was all these hydraulic parameters were generated from the hydrodynamic model for the same maximum discharge and variations in the value were due to variations of river physical shape.

## DISCUSSION AND CONCLUSIONS

In the present study the floodplain of lower Damodar river was classified into different zones based on high to low risk. This study incorporated the output of a hydrodynamic model along with topographical parameters for classifying the flood zone by ANN approach. A total of nine flood influencing factors, namely Water depth (m), Flow area (m<sup>2</sup>), Flow width (m), Velocity (m/s), Discharge (m<sup>3</sup>/s), Distance from river (m), Elevation(m), Slope pattern and River class were considered in an ANN model. Understanding the relative importance of flood-causing parameters is very important for classifying the flood zones. In the present study model performance was evaluated based on a confusion matrix. For taking care of unevenness and disproportionality in datasets, all evaluation metrics like precision, recall (Sensitivity), specificity and F1 score were calculated from the confusion matrix along with accuracy. Constant value of LR

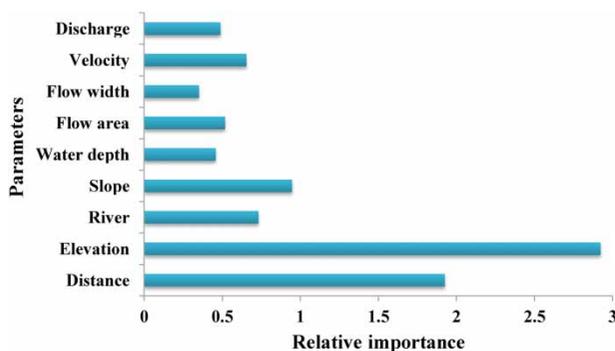


Figure 8 | Relative importance of parameters influencing flood zone.

i.e. 0.0001 was used during the training process and also to accelerate the algorithm iteration, and a momentum constant (Nesterov's momentum) of 0.9 value was used in the training process. The grid search technique was used for tuning the parameter during the training process. 60 hidden nodes were obtained for optimal performances by a grid search technique. In the present study the maximum number of iterations was terminated at 200 epochs. The model correctly predicted 16 sites into critical zone, 38 sites into high risk zone, 54 sites into moderate risk zone and 24 sites into low risk zone with an overall accuracy of 75.6%. Critical zones were identified by the model along all three streams namely Mundeswari, Amta Damodar and Amta Damodar branch and location of the critical zones was higher in number near Amta Damodar and Amta Damodar branch compared to Mundeswari river. The model predicts elevation and distance from the river as a higher influencing parameter for classifying the flood zone compared to other factors. Hence the site which was at very low elevation and very close to the river in topographical factors and showing high hydraulic factors namely discharge, velocity, flow width, water depth was classified as critical zone by the model. Influence of all the hydraulic parameter compared to topographical parameters as predicted by the present ANN model was less because the variations in the values of all the hydraulic parameters at different point of the cross sections of the same river were not very high. F1 score values for low zone, high zone, moderate zone and critical zone were 0.815, 0.818, 0.731 and 0.64 respectively. F1 score value shows that model prediction was on the higher side for low risk, moderate risk and high risk zone compared to critical zone. But the accuracy of critical zone was also reasonably good. This study can be further enhanced by employing best management practices (BMPs) in the flood prone region identified by the ANN model, which will be taken up in future research. The study can be used as a tool by decision managers to predict the risk of flooding in the study area with acceptable accuracy before flood events.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of IIT(ISM), Dhanbad for the FRS project (FRS (120)/2017-18/ME) of Prof. V.V. Govind Kumar for carrying out

this research work. Authors duly acknowledge Dept. of Civil Eng. and Dept. of Mining Eng. for providing the support and infrastructure facilities. The authors would like to thank DVC Maithon, CWC, Irrigation and Waterways Directorate, Govt. of West Bengal for providing the necessary data for this research study and DHI India for their technical support regarding the MIKE HYDRO RIVER model.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## REFERENCES

- Ahmad, S. & Simonovic, S. P. 2005 *An artificial neural network model for generating hydrograph from hydro-meteorological parameters*. *Journal of Hydrology* **315** (1–4), 236–251. <https://doi.org/10.1016/j.jhydrol.2005.03.032>.
- Ashrafi, M., Chua, L. H. C., Quek, C. & Qin, X. 2017 *A fully-online Neuro-Fuzzy model for flow forecasting in basins with limited data*. *Journal of Hydrology* **545**, 424–435. <https://doi.org/10.1016/j.jhydrol.2016.11.057>.
- Balica, S. F., Popescu, I., Beever, L. & Wright, N. G. 2013 *Parametric and physically based modelling techniques for flood risk and vulnerability assessment: a comparison*. *Environmental Modelling & Software* **41**, 84–92. <https://doi.org/10.1016/j.envsoft.2012.11.002>.
- Chang, L. C., Shen, H. Y. & Chang, F. J. 2014 *Regional flood inundation nowcast using hybrid SOM and dynamic neural networks*. *Journal of Hydrology* **519**, 476–489. <https://doi.org/10.1016/j.jhydrol.2014.07.036>.
- Chatterjee, C., Förster, S. & Bronstert, A. 2008 *Comparison of hydrodynamic models of different complexities to model floods with emergency storage areas*. *Hydrological Processes: An International Journal*, **22** (24), pp. 4695–4709. <https://doi.org/10.1002/hyp.7079>.
- Dawson, C. W. & Wilby, R. 1998 *An artificial neural network approach to rainfall-runoff modelling*. *Hydrological Sciences Journal* **43** (1), 47–66. <https://doi.org/10.1080/02626669809492102>.
- Dwarakish, G. S. & Ganasri, B. P. 2015 *Impact of land use change on hydrological systems: a review of current modeling approaches*. *Cogent Geoscience* **1** (1), 1115691. <https://doi.org/10.1080/23312041.2015.1115691>.
- Ghorbani, M. A., Zadeh, H. A., Isazadeh, M. & Terzi, O. 2016 *A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction*. *Environmental Earth Sciences* **75** (6), 476. <https://doi.org/10.1007/s12665-015-5096-x>.
- Haykin, S. 1999 *Neural Networks—A Comprehensive Foundation*. McMillan Inc, Englewood Cliffs, NJ, USA.
- Jahangir, M. H., Reineh, S. M. M. & Abolghasemi, M. 2019 *Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm*. *Weather and Climate Extremes* **25**, 100215. <https://doi.org/10.1016/j.wace.2019.100215>.
- Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A. & Moradi, A. 2012 *An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia*. *Environmental Earth Sciences* **67** (1), 251–264. <https://doi.org/10.1007/s12665-011-1504-z>.
- Kim, T., Heo, J. H. & Jeong, C. S. 2006 *Multireservoir system optimization in the Han River basin using multi-objective genetic algorithms*. *Hydrological Processes: An International Journal*, **20** (9), 2057–2075. <https://doi.org/10.1002/hyp.6047>.
- Kiş, Ö. 2007 *Streamflow forecasting using different artificial neural network algorithms*. *Journal of Hydrologic Engineering* **12** (5), 532–539. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:5\(532\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532)).
- Mosavi, A., Ozturk, P. & Chau, K. W. 2018 *Flood prediction using machine learning models: literature review*. *Water* **10** (11), 1536. <https://doi.org/10.3390/w10111536>.
- Nourani, V., Baghanam, A. H., Adamowski, J. & Kisi, O. 2014 *Applications of hybrid wavelet-artificial intelligence models in hydrology: a review*. *Journal of Hydrology* **514**, 358–377. <https://doi.org/10.1016/j.jhydrol.2014.03.057>.
- Pandey, M., Zakwan, M., Sharma, P. K. & Ahmad, Z. 2020 *Multiple linear regression and genetic algorithm approaches to predict temporal scour depth near circular pier in non-cohesive sediment*. *ISH Journal of Hydraulic Engineering* **26** (1), 96–103. <https://doi.org/10.1080/09715010.2018.1457455>.
- Prakash, O., Sudheer, K. P. & Srinivasan, K. 2014 *Improved higher lead time river flow forecasts using sequential neural network with error updating*. *Journal of Hydrology and Hydromechanics* **62** (1), 60–74. <https://doi.org/10.2478/johh-2014-0010>.
- Pramanik, N., Panda, R. K. & Sen, D. 2010 *One dimensional hydrodynamic modeling of river flow using DEM extracted river cross-sections*. *Water Resources Management* **24** (5), 835–852. <https://doi.org/10.1007/s11269-009-9474-6>.
- Rahmati, O. & Pourghasemi, H. R. 2017 *Identification of critical flood prone areas in data-scarce and ungauged regions: a comparison of three data mining models*. *Water Resources*

- Management* **31** (5), 1473–1487. <https://doi.org/10.1007/s11269-017-1589-6>.
- Shamseldin, A. Y. 2010 Artificial neural network model for river flow forecasting in a developing country. *Journal of Hydroinformatics* **12** (1), 22–35. <https://doi.org/10.2166/hydro.2010.027>.
- Singh, R. K., Villuri, V. G. K., Pasupuleti, S. & Nune, R. 2020 Hydrodynamic modeling for identifying flood vulnerability zones in lower Damodar river of eastern India. *Ain Shams Engineering Journal*. <https://doi.org/10.1016/j.asej.2020.01.011>.
- Sivapragasam, C., Maheswaran, R. & Venkatesh, V. 2008 Genetic programming approach for flood routing in natural channels. *Hydrological Processes: An International Journal* **22** (5), 623–628. <https://doi.org/10.1002/hyp.6628>.
- Smith, J. & Eli, R. N. 1995 Neural-network models of rainfall-runoff process. *Journal of Water Resources Planning and Management* **121** (6), 499–508. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1995\)121:6\(499\)](https://doi.org/10.1061/(ASCE)0733-9496(1995)121:6(499)).
- Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics* **10**, 3–22. <https://doi.org/10.2166/hydro.2008.015>.
- Solomatine, D. P. & Price, R. K. 2004 Innovative approaches to flood forecasting using data driven and hybrid modelling. In *Hydroinformatics* (in 2 volumes, with CD-ROM). pp. 1639–1646. [https://doi.org/10.1142/9789812702838\\_0202](https://doi.org/10.1142/9789812702838_0202).
- Tehrany, M. S., Pradhan, B. & Jebur, M. N. 2015 Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stochastic Environmental Research and Risk Assessment* **29** (4), 1149–1165. <https://doi.org/10.1007/s00477-015-1021-9>.
- Wardlaw, R. & Sharif, M. 1999 Evaluation of genetic algorithms for optimal reservoir system operation. *Journal of water resources planning and management* **125** (1), 25–33. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1999\)125:1\(25\)](https://doi.org/10.1061/(ASCE)0733-9496(1999)125:1(25)).
- Yamazaki, D., Kanae, S., Kim, H. & Oki, T. 2011 A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, **47** (4). <https://doi.org/10.1029/2010WR009726>.
- Young, C. C., Liu, W. C. & Wu, M. C. 2017 A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Applied Soft Computing* **53**, 205–216. <https://doi.org/10.1016/j.asoc.2016.12.052>.
- Yu, P. S., Yang, T. C., Chen, S. Y., Kuo, C. M. & Tseng, H. W. 2017 Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology* **552**, 92–104. <https://doi.org/10.1016/j.jhydrol.2017.06.020>.

First received 14 July 2020; accepted in revised form 25 September 2020. Available online 14 October 2020