Optimization of state-of-the-art fuzzy-metaheuristic ANFIS based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India

Aman Arora¹, Alireza Arabameri², Manish Pandey³, Masood A. Siddiqui¹, U.K. Shukla⁵, Dieu Tien Bui⁶, Varun Narayan Mishra⁷, Anshuman Bhardwaj⁸

¹Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi-110025
²Department of Geomorphology, Tarbiat Modares University, Jalal Ale Ahmad Highway, Tehran 9821, Iran
³University Center for Research & Development (UCRD), Chandigarh University, Mohali-140413, Punjab, India
⁴Department of Civil Engineering, Chandigarh University, Mohali-140413, Punjab, India
⁵Center for Advanced Study in Geology, Institute of Science, Banaras Hindu University, Varanasi-221005, India
⁶Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
⁷Centre for Climate Change and Water Research, Suresh Gyan Vihar University, Jaipur-302017, Rajasthan, India
⁸School of Geosciences, University of Aberdeen, Meston Building, King’s College, Aberdeen, AB24 3UE, UK

*Corresponding author: manish07sep@gmail.com; aman.jmi01@gmail.com
Abstract

This study is an attempt to quantitatively test and compare novel advanced-machine learning algorithms in terms of their performance in achieving the goal of zonation of predicting flood susceptible areas in a low altitudinal range, sub-tropical floodplain environmental setting like that prevailing in the Middle Ganga Plain (MGP), India. This part of the Ganga floodplain region, which under the influence of undergoing active tectonic regime related subsidence, is the hotbed of annual flood disaster, and is one of the best natural laboratories to test the flood susceptibility models for establishing a universalization of such models in low relief highly flood prone areas. Based on highly sophisticated flood inventory archived for this region, and 12 flood conditioning factors viz. annual rainfall, soil type, stream density, distance from stream, distance from road, Topographic Wetness Index (TWI), altitude, slope aspect, slope, curvature, land use/land cover, and geomorphology, an advanced novel hybrid model Adaptive Neuro Fuzzy Inference System (ANFIS), and three metaheuristic models-based ensembles with ANFIS namely ANFIS-GA (Genetic Algorithm), ANFIS-DE (Differential Evolution), and ANFIS-PSO (Particle Swarm Optimization), have been applied for zonation of the flood susceptible areas. The flood inventory dataset, prepared by satellite based collected flood samples, were apportioned into 70:30 classes to prepare training and validation datasets. One independent validation method, the Area-Under Receiver Operating Characteristic (AUROC) Curve, and other 11 cut-off-dependent model evaluation metrics have helped to conclude that the ANIFS-GA has outhustled other three models with highest success rate AUC = 0.922 and prediction rate AUC = 0.924. The accuracy was also found highest for ANFIS-GA during training (0.886) & validation (0.883). Better performance of ANIFS-GA than the individual models as well as some
ensemble models suggests and warrants further study in this topoclimatic environment using other classes of susceptibility models for the purpose of establishing a benchmark model with capability of highest accuracy and sensitivity performance in the similar topographic and climatic setting taking assumption of the quality of input parameters as constant.

**Keyword:** Flood Susceptibility Mapping, ANFIS, Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO), metaheuristic optimization, Middle Ganga Plain

1. Introduction

Floods, one of the most pervasive natural phenomena, are inflicting life, property (including livestock), soil loses (Keesstra et al., 2018), biodiversity losses (Keesstra et al., 2016) and ecological losses (Barredo, 2009; Cai et al., 2011; Yevjevich, 1994). Climate change induced flooding trend and pattern are increasing the complexity of this phenomena and related damages in the coastal areas is predicted to increase in the impending future (Hallegatte et al., 2013; Hanson et al., 2011). The flood related damages cannot be ignored and the complete prevention from it is unfeasible. It is this damage potential of floods that the United Nations has included this issue in its “United Nations Sustainable Development Goals (UNSDGs)” to properly tackle this menace and losses therefrom in a well-planned and strategic manner (UNSDG, 2013). The complex nature of floods, immensity of flood losses, and inclusion of flooding problem in the UNSDGs has made it essential and urgent to scientists to propose desirable and more accurate flood prediction methods for helping hazard management and mitigation agencies in better preparedness, planning for operational easiness, and make effective and adequate mitigation plans to curtail the damages from future flood events. The present flood modelling community is
attempting to develop and adopt more rigorous and logical mathematical approach to
delineate the flood susceptible regions at different spatial scales including large and
small scales. The approach inculcates various steps, some of the momentous ones
are: preparation of inventory data with reasonable accuracy for training and
validation purposes, the selection of most significant and high potential flood
conditioning or geo-environmental factors which are suitable to tectonoclimatic and
regional setting of the study area, digital elevation model (DEM) and satellite data
with good spatial resolution to prepare the conditioning factor datasets, use of
accurate & appropriate models, and assessment of the performance of models using
advanced and reliable evaluation techniques. Among these steps, the step wherein
the flood inventory data is prepared, utmost care of its accuracy must be taken into
consideration because the training of models mostly depends on the quality of the
inventory datasets; also, the accuracy of flood susceptible maps is compromised if
the flood inventory data quality is not good (Khosravi et al., 2016a; Merz et al., 2007).
The topography and hydro-climatological conditions influence the flood directly or
indirectly, therefore, the selection of the flood controlling factors should also be done
taking the type of topography under consideration. From an exhaustive literature
survey and experts’ knowledge of the field conditions, a list of common factors that
are selected for spatial modelling have been prepared which includes altitude,
aspect, slope, curvature, rainfall, soil, land use, land cover, stream density, distance
to stream or rivers, distance to road, and topographic wetness index (TWI)
(Arabameri et al., 2019e; Chapi et al., 2017a; Khosravi et al., 2016b; Sachdeva et
al., 2017; Samanta et al., 2018; Tehrany et al., 2013). In recent studies the selection
of additional factors is observed as normalized difference vegetation index (NDVI),
topographic positional index, topographic ruggedness index (TRI), stream potential
index (SPI), and various other proxy factors that are prepared based on topography or/and hydro-climatology of the region. In hilly region, where the flash-floods are common occurrence, the nature of most important flood contributing factors’ significance levels changes. In low altitudinal floodplain settings characterized by humid tropical to subtropical climate, geomorphology plays important role as compared to lithology. Whereas in mountainous watersheds, slopes and other factors play a relatively more significant role in flood occurrence potential prediction (Arora et al., 2019). Therefore, a caution should be practiced when selecting flood controlling factors for low altitude range viz. floodplain environment infested with conditions conducive to annual riverine floods and mere careless selection of flood control factors from the published literature irrespective of their topographic and climatic setting may cause diminished model performance. The study by Arora et. al. (2019) was conducted in the same area, by applying the level of caution as per the above suggestions, including a new control factor that is detailed microscale geomorphology, which has never been incorporated in previous studies of flood susceptibility prediction and had found that the geomorphology had not only played a better role than geology for flood susceptibility mapping for floodplains but its significance level was found to excel all other conditioning factors included in that study. In the earlier analysis, for the same study area, two models, one bivariate and one machine learning, were chosen and the prediction rates were found more than 80% for Frequency Ratio & Shannon’s Entropy models. In present study, the geomorphology dataset has been upgraded and the previous geomorphology dataset was replaced by this new dataset with an aim to achieve different and better performance result.
The study area is the part of the Ganga river basin (GRB) where the monsoon induced heavy rainfall related annual flooding is a common phenomenon. The study area encompasses the confluences of major rivers flowing over the flood plain of GRB and also shares the state boundaries of Uttar Pradesh and Bihar. The majority of floods are reported by an international agency, Emergency database (EM-DAT), during 1985-2015 in states of Bihar, Uttar Pradesh, Assam, West Bengal, and Orissa in India (CRED, 2016). In most of the states where flood occurrences are very high, the prevailing topographic settings are conditioned with plain, low relief topography and lithological variability is minimal as the vast floodplains of the Ganga River and its tributaries are the source of sediment forming the lithological units, the prioritized selection of geomorphology over geology dataset appears to be logically strong and valid choice to test performance of various models of the flood susceptibility mapping (FSM) in such alluvial plain landscape. Again, whether the selected factors have interdependence on one another or have collinearity problem can be checked with multicollinearity analysis wherein the algorithm checks the collinearity among conditioning factors (Alin, 2010; Chen et al., 2018; Costache et al., 2020a). Afterwards, the relative importance of each factor can also be analyzed by different methods before applying models for spatial susceptibility prediction modelling; among them the frequently used methods are information gain method (Costache and Tien Bui, 2019; Đurić et al., 2019), Binary logistic regression (Arabameri et al., 2019b), Boosted regression tree (Arabameri et al., 2019f; Zabihi et al., 2019), and random forest model (Arabameri et al., 2019c). Under prevailing analysis the random forest model has been used to assess the competency of conditioning factors to fulfill the requirement for spatial modelling applications-based analysis. In order to achieve greater benefit from the conditioning factors, the selection of the satellite data or
other existing sources, data/maps, should also be of commendable quality to prepare these datasets. Among them, the hydrological and topographical based conditioning factor datasets are prepared through the remotely sensed imagery data and digital elevation model (DEM) data. A better spatial resolution remotely sensed satellite imageries also improve the quality of hydrological and topographical based conditioning factor datasets. Many studies have been including the ALOS PALSAR DEM dataset (Arabameri et al., 2019c), in place of lesser resolution shuttle radar thematic mapping (SRTM) DEM for preparation of topographic based conditioning factors. It may be noted that the ALOS PALSAR DEM the 12.5-meter spatial resolution dataset is freely available DEM data in public domain which is better than the SRTM DEM resolution of 30 meters for mountainous fluvial landscape studies (Boulton and Stokes, 2018). But in low altitude range environments SRTM 30m version 3 is vindicated to be widely acceptable in hydrological and landscape studies (Hayakawa et al., 2008; Zhang et al., 2019).

From last two decades, there have been continuous studies conducting FSM using different statistical techniques; it varies from bivariate statistical methods to multivariate models. The current trend in the natural hazard studies of large regions is focused on the use of machine learning (ML) (Achour and Pourghasemi, 2020; Hæring et al., 2012; Hong et al., 2018b; Tehrany et al., 2015; Wang et al., 2020), multi-criteria decision making (MCDM), (Costache et al., 2020c; Oh et al., 2018; Santos et al., 2018) and ensembles models, combining two or more models, using bivariate, ML, and MCDM category models. There is no restriction on the selection of models for developing the ensembles; in other words, the selection of one model for creating the ensemble can be done from two or more different categories of models. It has also been observed in the flood modeling literature that the ensemble models
have performed more accurately than individual or standalone models (Bui et al., 2019; Chapi et al., 2017b; Choubin et al., 2019a; Shahabi et al., 2020). Among these published works, a study for FSM carried out by Chapi et al., (2017b) in Haraz watershed in northern Iran encompassing 4014 km\(^2\) study area shows that the combination of bagging ensemble with Logistic Model Tree (LMT) model has performed better than the LMT, logistic regression (LR), Bayesian logistic regression (BLR), and random forest (RF) standalone models. In another study, Tehrany et al., (2015) had attempted Support Vector Machine (SVM) based ensembles prepared by using different kernel types to predict the flood susceptible regions, and their results demonstrated that SVM based ensembles have much better prediction rate, >81%, than the standalone frequency ratio (FR) model. Recently, another class of hybrid model combining Adaptive Neuro Fuzzy Inference System (ANFIS) with Artificial Neural Network (ANN) and Fuzzy Logic models, has emerged to be frequently used for prediction of flood susceptible areas (Ahmadlou et al., 2019; Bui et al., 2018b; Hong et al., 2018b; Razavi Termeh et al., 2018). For instance, the study by Bui et al., (2018a) used ensembles combining the ANFIS with three metaheuristic models for FSM for a study area located in Iran and found that all three models have surpassed 94% prediction rate accuracy level. This study too, has vindicated the supremacy of ANFIS based ensembles compared to the standalones used in the study.

Benchmarking the flood susceptibility model, that is establishing the best performing most accurate and highly sensitive flood prediction model for a specific type of topoclimatic environment, keeping the quality of input parameters and model configuration settings constant, requires testing and validating all type of available models in each and every topoclimatic settings. It should be noted that there is considerable number of research work published in different parts of the world by
using the ANFIS based FSM, but to date, no such study applying this category of
advanced novel flood ensemble models has been conducted for Ganga River Basin
(GB, India or in any part of India or in any part of the world with the such low
altitudinal range topographic setting with characteristic climatic setting as MGP. In
MGP, the southwest monsoon led heavy rainfall brings havoc every year, especially
in lower basins of the GRB, and debouches heavy rainwater in the entire catchment
and causes substantial amount of damages. Therefore, there is an immediate
requirement in pursuit of benchmarking of flood susceptible models for this part of
the Ganga river basin wherein some bivariate and machine learning individual
models have already been tested and reported. This study is the next step in the
pursuit of susceptibility model benchmarking in low relief sub-humid topoclimatic
environmental setting and such study with other available highly advanced recent
models must continue until the best model out of all the available models is
achieved.

Keeping in view the research gaps and orientation of research of flood susceptibility
modeling community, the aim of this study has been set to assess and analyze the
performance of novel advanced models, ANFIS (standalone) and three of its hybrids
optimized with Genetic Algorithm (GA), Differential Evolution (DE), and Particle
Swarm Optimization (PSO) viz. ANFIS-GA, ANFIS-DE, ANFIS-PSO, for prediction of
flood susceptible zones in the MGP. The novelties of this study lie in the first time
use of ANFIS and three of its advanced ensembles for FSM and evaluation of the
different aspect of models’ performances through the use of a hot of cut-off-
dependent & cut-off-independent matrices.

2. Description of the study area
The Middle Ganga Plain (MGP) is the most densely populated part of the Ganga River Basin (GRB) (Singh, 1971). And, the present study area holds the densest populated region of MGP. The study area shares the most densely populous districts, Patna, Siwan, Saran, Vaishali, Samastipur, Begusarai, Nalanda, and Bhojpur of Bihar state. Geographically, the study area extends between 25°10’N-26°10’N and 83°90’-85°45’E latitudes and longitudes. It accounts for a total of 10,138 km² area of upper and lower GRBs. The MGP witnesses floods almost every year during monsoon and even the post-monsoon seasons do not remain dry and cause flooding under the influence of the South-West monsoon rainfall (Vittal et al., 2016).

In the MGP, there are confluence points for five major rivers and their tributaries viz. Ganga, Gandak, Ghaghara, Kosi, and Son (Arora et al., 2019). Therefore, such vast upper catchment area with latitudinal climatic and topographic gradient enhances the proneness of these low lying floodplains, having major confluence zones of big rivers, to severe flooding every year during both monsoon as well as post-monsoon periods.

The region experiences summer (April–May), monsoon (June–September), post-monsoon (October–December) and winter (January to March) seasons in a year (Dimri et al., 2019). Due to sub-tropical-humid climate the MGP observes the highest temperature during the months from April to July. It ranges between 35 to 45°C during summer season. The lowest temperature 3-4°C is observed during the months of December and January. The southwest monsoon hits the MGP during sometime between late June and early July. The high intensity rainfall in the upper catchments of GRB causes floods in the MGP in post monsoon season, also, during the months of August and September. The in-situ water discharge data, for Gandhighat station, Patna, obtained from Central Water Commission (CWC), Govt.
of India, reveals that the maximum average monthly water discharge is recorded
during the months of August & September which result into the post monsoon
flooding in the MGP. From the collected water discharge data, it has been observed
that the average annual discharge for Gandhighat station soars up to 50,000m$^3$/sec
in a single day, during the flood months. The high water discharge in the streams
increases the possibility of levee breaches and bank overflow resulting into flooding.
The location map of CMGM overlapped with district-wise no. of household and the
recorded average monthly discharge of Ganga river at Gandhighat station during
monsoon period (June to October) for year 2008 is demonstrated in figure 1.

3. Materials and methods

3.1 Data used

3.1.1 Flood inventory data

The flood inventory data should be prepared with accuracy before using it for
extracting random sample points to be later used of model training and validation in
the studies related to flood (Choubin et al., 2019b), whereas, precision in the sample
point collection process increase model accuracy for prediction (Arora et al., 2019;
Tehrany et al., 2013). For generation of accurate flood inventory, the Landsat
satellite data, series 5, has been downloaded from the Earth explorer website
(https://earthexplorer.usgs.gov/) for the pre-flood, during-flood, and post-flood
periods. The details of the downloaded datasets are provided in table 1. The
complete exercise of preprocessing of satellite imageries and retrieval of inundated
pixels using Normalized Difference Water Index (NDWI) has been already explained
by Arora et al. (2019).
In present work, we have generated a total 1000 random floodpoints and the same number of non-inundation points have been also generated. Out of which, 70% have been utilized for training and remaining 30% have used for testing purposes.

3.1.2 Conditioning factors (CgFs)

The construction of CgFs datasets was accomplished by using the SRTM 1-arc sec (30m) DEM along with re-analysed meteorological data, and some existing stream maps. The DEM generated CgFs are Altitude, Curvature, Distance from streams, Stream density, Slope (degree), Slope Aspect, and TWI. The Land Use Land Cover (LULC) map for the year of 2008 was downloaded from the European Space Agency (ESA) led Climate Change Initiative (CCI) (https://www.esa-landcover-cci.org/). The ESA has archives of generated annual LULC data for the period 1992-2015 at global scale at 300m resolution and provides the same on their web portal for utilization in studies related to non-commercial purposes (Li et al., 2018). The soil data was procured from the Food and Agriculture Organization (http://www.fao.org) of the United Nations. The geomorphology data was procured in vector format from the Bhukosh Portal of the Geological Survey of India website (http://bhukosh.gsi.gov.in/Bhukosh/Public). The rainfall data was prepared by the retrieved data from Climate Forecast System Reanalysis (CFSR) web-portal (https://globalweather.tamu.edu/). The distance to road data was prepared by using the downloaded road network from Open Street Map portal (https://www.openstreetmap.org/). Thereafter, to achieve the homogeneity, all collected data was converted into raster format and resampled at 30-meter resolution to use further for spatial modelling.
The altitude range in the study area is so low (13 - 96 m) that it appears almost homogenously featureless topographic setting when viewed in medium resolution dataset. This is the reason why the vast floodplains of GRB have long been considered featureless (Singh, 1971) and still there is no geomorphological maps at finer scales like 1:25000 or 1:5000. The altitude dataset is demonstrated with training and validation sample points in figure 3A.

Due to plain topography, the plan curvature in most parts of the MGP is flat type. The convex and concave types of curvature can mainly be found along river banks or along major topographic break points like terraces etc. (figure 3B) The figure 3C & figure 3D are showing the distance to road & distance from stream datasets, respectively. Both of the datasets have been prepared using Euclidean distance analysis method. The maximum distance of 7073 meters was noted for the road network and 5383 meters for the stream network from their respective connected points. The equation (1) of Euclidean distance method is mathematically expressed as follows:

\[ d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  

Where \( x \) and \( y \) are two points in Euclidean n-space and \( d \) refer distance between \( x \) and \( y \).

The geomorphology map of the MGP is displayed in figure 3E. The geomorphology dataset is combined form of 21 geomorphological units; each unit plays a significant role in flood prediction potential. The most important unit is the FluOri- active flood plain unit where the probability of the flood occurrences is very high. The detailed information of each unit is provided in the table 2.
LULC is the mosaic of all the natural and man-made features; both categories, of the earth surface, have influence over hydro-meteorological processes, directly and indirectly. The LULC map (figure 3F) is categorised into four categories- Agricultural Lands, Vegetation/Treecover, Built-up Lands, and Water Body (Rivers/Ponds). The maximum number of flood sample point is observed to be in agricultural lands class of LULC.

The rainfall dataset prepared using the interpolation analysis of collected CFSR data for ~36 years (Jan 1979-Jul 2014). The numerical average data for all stations were assembled before running the Inverse Distance Weighted (IDW) interpolation algorithm (equation 2). The IDW is widely used in earth sciences related studies (Bartier and Keller, 1996). The prepared rainfall map, displayed in figure 3G, provides the spatial overview of rainfall distribution. The highest average annual rainfall was recorded 1281mm for MGP whereas the lowest value was noted to be 1002mm only. The highest rainfall is observed in south-west and south-east parts of MGP represented by dark blue in the displayed map.

$$Z_{x,y} = \frac{\sum_{i=1}^{n} z_i w_i}{\sum_{i=1}^{n} w_i} \quad \text{(2)}$$

Where $Z_{x,y}$ refers to the point value (rainfall) to be estimated, $z_i$ represents control value for $i_{th}$ sample point, and $w_i$ is a weight that determines the relative importance of individual control point $z_i$ during interpolation process.

Slope aspect indirectly influence the floods through employing control on various geo-environmental factors i.e. soils, vegetation, rainfall, etc. (Rahmati et al., 2016). The slope aspect map (figure 3H) has been categorized in nine divisions ranging...
from flat to Northwest zones. The following equation (3) was used to calculate the slope aspect:

\[
\text{Aspect} = 270^\circ + \arctan \left( \frac{f_y}{f_x} \right) - 90^\circ \frac{f_x}{|f_y|}
\]  

(3)

Where, \( f_x \) and \( f_y \) are the rate of change in latitude values measured in the north-south and the east-west directions, respectively (Hu, 2016).

The slope of a region controls and influences the hydrological activities; affects the surface directly (Tehrany et al., 2019). Due to the plain topography a very low variance in the slope for MGP existed. The slope map (figure 3I) provides the detailed visualization of slope (in degree) for the study area.

In MGP, total six soil types are found as shown in soil map (figure 3J) of MGP. Among all six categories of soils, the maximum spatial coverage is observed by the FL-Fluvisols (3743) category. The FL-Fluvisols category covers the active floodplain mostly as well as it covers the maximum share of flood samples.

Per unit area of the total length of all stream networks is called the stream density or drainage density (Sangireddy et al., 2016). The probability of flood occurrences is very high at high stream density places (Onušluel Gül, 2013). The stream density map (figure 3K) prepared by using the line density tool of ArcGIS toolbox. The maximum and minimum density was recorded 75.60 km/km\(^2\)& 0.96 km/km\(^2\) respectively.

The TWI is calculated by dividing the specific basin area by the slope of the region. The TWI map is displayed in figure 3L. The mathematical representation, equation (4), of TWI calculation is as follows:
Where $A_s$ represents the particular catchment area and $\beta$ indicates slope angle.

### 3.2 Methodology

The flowchart of methodology is provided in figure 2. The following section would provide a brief detail of models used in this study.

#### 3.2.1 Description of models

#### 3.2.1.1 Multi-Collinearity Analysis

The multicollinearity analysis (MCA) helps to find collinearity among conditioning factors (CgFs), if present, and allow a user to decide the non-collinear CgFs for further analysis in order to get unbiased results by spatial modelling. Hence, the MCA is recommended for a spatial modelling study (Arora et al., 2019). The published research works suggest use of two indices, Variance Inflation Factor (VIF) and Tolerance (TOL) under MCA to evaluate collinearity dependency of CgFs among each other and, subsequently, the CgFs found suitable for further analysis if the TOL > 0.1 & VIF < 10 retrieved in results (Hair et al., 2013; Menard, 2002).

These two indices, TOL and VIF, can be calculated by the following equations:

\[
TOL = 1 - R^2_v \\
VIF = \frac{1}{TOL}
\]

Where $R^2_v$ represents the ‘coefficient of CgFs on all other CgFs’

#### 3.3.2 Determination of relative weights for conditioning factors

\[
TWI = \ln \left( \frac{A_s}{\beta} \right)
\]
The Random Forest (RF) model is an ensemble-based learning algorithm where a large pool of decision-trees (DTs) are constructed to perform the spatial relationships between an event (occurrence of flood) and the related factors (CgFs) for classification (Lee et al., 2017). A pair of decision trees analyses the classification (also refers as ‘vote’ in model) and the maximum number of votes by the class concluded as the results of RF (Arabameri et al., 2019d). The determination of each factor is decided by the importance of used variables in the model. In comparison to other decision tree methods the results generated by RF model found with less errors due to the use of multitude set of trees. At this instance, the RF model has been used to determine the relative weights or importance of each CgFs.

The flood conditioning datasets $FCD = (x_i, y_i)_{i=1}^N$ with $x_i \in R^M$ is the CgFs; where ‘$N$’ refers the total number of samples and ‘$M$’ is the total count of CgFs. $y_i \in (1,0)$ is the output which contains flood (1) & non-flood (0) occurrences. In first step, the bootstrap algorithm used to generate $n$ subsets (bootstrap subsets), where each subset consists of ‘$m$’ factors; where $m \leq M$. Then, in second step, the CART (Classification And Regression Tree) algorithm has been used to construct the tree classifier for each bootstrap subset (Arabameri et al., 2019f). In final stage, all constructed classifiers have been aggregated to design a RF classifier. In general, to construct a RF model, two parameters: ‘$n$’ & ‘$m$’ should be determined, where ‘$n$’ should be large enough, i.e. ‘$n$’= 500 (Tien Bui et al., 2016) to contrast the RF model. The total number of variables was restricted to 3 at each split of the model and the Out-of-bag (OOB) estimate of error rate was found 30.34% for the performed model.

3.3.3 Flood susceptibility mapping (FSM)
The FSM has been conducted in the study through four models are- ANFIS and ensembles with GA, DE, & PSO heuristic models. Brief information of all four models has been provided in following sub-sections.

3.3.3.1 Adaptive neuro-fuzzy inference system (ANFIS)

The neuro-fuzzy network, ANFIS, was proposed in 1993 (Jang, 1993) and since then is being used in different disciplines of sciences including geosciences (Ahmadlou et al., 2019; Lei et al., 2007; Naderloo et al., 2012; Najafi and Faizollahzadeh Ardabili, 2018; Razavi Termeh et al., 2018; Republic, n.d.; Wei et al., 2007). The output of this adaptive network depends upon the parameters of nodes and to optimize the performance during training of the parameters there are two kinds of learning algorithm applied to tune these parameters (Li and Su, 2010). In other words, the system of ANFIS can be understand as there are two inputs, x & y, and an output z. There are five layers or steps, presented by if-then fuzzy rules, based on Sugeno model, to calculate the output (Takagi and Sugeno, 1985).

\[ \text{Rule 1: if } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{then } f_1 = p_1x + q_1y + r_1 \]  \hspace{1cm} (7)

\[ \text{Rule 2: if } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{then } f_2 = p_2x + q_2y + r_2 \]  \hspace{1cm} (8)

Where the fuzzy sets are \( A_1, A_2, B_1, \) and \( B_2 \) and the modifiable parameters are represented by \( p_i, q_i, \) and \( r_i \); \( f_i \) indicates separated outputs for both rules.

Layer 1: every node \( i \) is an adaptive node with node function

\[ O_i^1 = \mu A_i(x), i = 1,2 \]  \hspace{1cm} (9)

or
Where \( x \) (or \( y \)) is the input to the node \( i \) and \( A_i \) (or \( B_{i-2} \)) represents linguistic label associated with node. \( O_i^1 \) or \( O_i^1 \) are the membership function for \( A_i \) (or \( B_{i-2} \)), specifies the degree where the given \( x \) or \( y \) satisfies \( A_i \) or \( B_{i-2} \). The \( A \) and \( B \), the membership functions, usually denote as bell functions (Übeyli et al., 2010):

Generally, \( \mu A_i(x) \) is chosen to be bell-shaped with maximum and minimum equal to 1 and equal to 0 respectively, such as

\[
\mu A_i(x) = \frac{1}{1 + \left| \frac{x - r_i}{p_i} \right|^{2q_i}}
\]

where \( \{p_i, q_i, r_i\} \) are called premise parameters.

Any continuous and piece-wise distinguishable function, like a triangular-shaped membership function, is also a qualified member for the node function in this layer (Catalão et al., 2011). The parameters in this layer are noted as premise parameters.

Layer 2: Every node is a fixed node, represents the firing strength of the rule, in this layer, and acts multiplies the incoming signals/parameters. The output represents as;

\[
O_i^2 = w_i = \mu A_i(x) \times \mu B_i(y), \ i = 1,2
\]

Layer 3: In this layer, each node refers an adaptive node labelled as ‘N’, computes the ratio of the \( i_{th} \) rule’s firing strength to the sum of all rules’ firing strengths:

\[
O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \ i = 1,2
\]

Generally, the outputs are labelled as normalized firing strengths.
Layer 4: In this layer, each node called an adaptive node, computes the contribution of the $i_{th}$ rule to the overall output, with a function:

$$O_i^4 = \bar{w}_i z_i = \bar{w}_i (a_i x + b_i y + c_i), \quad i = 1, 2$$

(14)

Where, $\bar{w}_i$ indicates the output of layer 3, and $(a_i, b_i, c_i)$ are the parameter set; also referred as consequent parameters.

Layer 5: In final layer or layer 5, the single node or fixed node calculates the overall output by summing all incoming/retrieved signals from preceding ones:

$$O_i^5 = \sum_i \bar{w}_i z_i = \frac{\sum_i w_i z_i}{\sum_i w_i}, \quad i = 1, 2$$

(15)

3.3.3.2 Genetic algorithm (GA)

The GA was first proposed in 1989 (Booker et al., 1989) by inspiring from the Charles Darwin's theory: 'survival of the fittest' (McCall, 2005). The basic concept behind the robust GA is to perform natural selection test, even in noisy environments. The algorithm discovers the best output or solution among resultant outputs in three steps- 'Selection', 'Crossover, and 'Mutation'.

In each step, the algorithm randomly selects and analyse individuals from the population and further utilize them to produce the new data or children for the next generation and it repeats until retrieve the ideal result (Li and Su, 2010).

3.3.3.3 Differential evolution (DE) algorithm

The DE, a stochastic, population-based optimisation and an evolutionary algorithm developed to finding a solution (Price et al., 2006; Storn, 1999). Here, a brief detail of
the algorithm is explained, the complete details may read elsewhere (Storn and Price, 1995). The algorithm performed in four steps, Initiation, Mutation, Crossover, and Selection, to solve a complex problem.

*Initiation:* In initiation, the initial population, $N_p$ size, randomly generated; where, the decision parameters for each values are assigned (Tien Bui et al., 2017).

\[ X_{j,i} = X_{j}^{min} + m_j(X_{j}^{max} - X_{j}^{min}) \]  
(16)

Where $i = 1, 2, ..., N_p$ and $j = 1, 2, ..., D$; [$D$ indicates decision variables number, $X_{j}^{max}$ is upper bound while $X_{j}^{min}$ is indicating the lower bounds of the decision $j^{th}$; and ‘m’ is a random number in ‘0’ to ‘1’ range.

*Mutation:* The mutation operator generates mutated or differential (donor) vectors in which each vector constructed by following rule;

\[ X_{j,gen+1} = X_{r1,gen} + F(X_{r2,gen} - X_{r3,gen}) \]  
(17)

where $r1$, $r2$, and $r3$ are randomly chosen indices [$r1 \neq r2 \neq r3 \neq i$]; $X_{r1,gen}$, $X_{r2,gen}$, and $X_{r3,gen}$ are randomly selected valued from the populations, and $F$ is the scale factor.

*Crossover:* This operator used donor & target vectors to produce a trial vector(Hong et al., 2018a);

\[ X'_{j,i,gen+1} = \begin{cases} X_{j,i,gen+1} & \text{if } \rho_j \leq C_R \text{ or } j = q \\ X_{j,i,gen} & \text{else} \end{cases} \]  
(18)

Where $\rho_j \in [0, 1]$ is a uniform random number, $C_R \in [0, 1]$ represents the crossover parameter, and $q \in \{1, ..., D\}$ represents as an index as assurance for selection of one of the parameter, at least, in the mutant vector.
Selection: In comparison of fitness value of trial vector with target vector, the selection operation is used to choose the best choice for the next generation;

\[
X_{i,gen+1} = \begin{cases} 
X'_{i,gen+1} & \text{if } f(X'_{i,gen+1}) \leq f(X_{i,gen}) \\
X_{i,gen} & \text{else} 
\end{cases} \tag{19}
\]

Where, \(X_{i,gen}\) represents parent vector, and \(f(*)\) is called the fitness function.

3.3.3.4 Particle swarm optimization (PSO) algorithm

The swarm intelligence, inspired by the shared behaviours of animal was used to find out the distributed solutions (Chen et al., 2020). The PSO algorithm, based on swarm intelligence, was developed by Kennedy and Eberhart, (1995) to know the social behaviour of birds flock and their food catching processes. In some aspects, PSO can be referred as an evolutionary method such as GA. But in PSO the genetic operation is unavailable, which is used to produce a next generation solution in GA.

In PSO algorithm, a random number of feasible solutions, described by vectors, generated at the time of initialization of it to find the optimal solution (Mehrabi et al., 2020). The solution in PSO is also referred as particle. In order to evaluated the potential of each particle, the velocity has been assigned to each one of them, who have been positioned randomly, by calculating the fitness value \(p_{best}\) and the elite position \(g_{best}\) of each particle (Moayedi et al., 2020). The PSO algorithm process can be expressed mathematically as;

\[
N'_i = wN_i + C_1r_1(p_{best} - R_i) + C_2r_2(g_{best} - R_i) \tag{20}
\]

\[
R'_i = R_i + N'_i \tag{21}
\]

The computation to change the velocity \(R_i\) of each particle toward the location of the identified \(p_{best}\) and \(g_{best}\).
Where, $N_i$ represents the particle position; $C_1$ and $C_2$ represent the cognitive and social scaling parameters respectively, and $R_i$ is the velocity of each particle. The $r_1$ and $r_2$ symbolize the random numbers between 0 and 1 and $w$ is the inertia weight.

### 3.4 Evaluation and Comparison of models

The performance assessment of involved models has been completed by two methods: cut-off-independent and cut-off-dependent methods. The receiver operating characteristics (ROC) as a cut-off-independent evaluation method is found most reliable and successful robust technique to assess the model performance (Arora et al., 2019; Schumann et al., 2014). Whereas, the dependent matrices evaluation methods, i.e. accuracy, F-score, sensitivity, specificity, odd ratio, and Cohen’s Kappa, etc., have been utilized to assess the model performance along with ROC to have evaluation of different facets of model performances. The precise model evaluation requires both dependent and independent matrices evaluation methods (Rahmati et al., 2019). Rahmati et al. (2019) have reviewed 21 cut-off-dependent matrices and have defined clearly what those matrices refer to. These different cut-off-indices may be of use to end user agencies interested different aspects of the model performance. For example, if the end user agency expresses desire to know a model’s ability to incorrectly predict non-flood events, it will directly look on that models, FPR (or fall out) which equals ‘1-specificity). Similarly, if the agencies’ need is to know overall error rate, it will look into the ‘misclassification rate’ column of cut-off-dependent indices table.

In order to calculate the results based on above discussed evaluation methods, the confusion matrix for training and validation datasets for flood modelling in $2 \times 2$ row and column format table is prepared. Where, four types of possible consequences,
including true positive (TP), false positive (FP), true negative (TN), and true positive (TP), occurrences were analysed. The TP indicates the correctly classified pixels of flood event; FP is representing the count of incorrectly classified pixels as flood event. Whereas, the correctly classified pixel as non-flood is called TN and FN describes the number of incorrectly pixels classified as non-flood. Based on these four possible consequences, TP, FP, TN, & FN, the specificity, true positive rate (TPR), sensitivity, false positive rate (FPR), false discovery rate (FDR), false negative rate (FNR), accuracy, precision, F-Score, accuracy, odd ratio and Cohen’s kappa statistics are calculated and equations of these metrics are as follows:

\[ TPR = Sensitivity = \frac{TP}{(TP+FN)} \]  
\[ \text{Specificity} = \frac{TN}{(TN+FP)} \]  
\[ FPR = \frac{FP}{(TN+FP)} = (1 - \text{Specificity}) \]  
\[ FDR = \frac{FP}{(TP+FP)} \]  
\[ FPR = \frac{FP}{(FN+TP)} \]  
\[ \text{Accuracy} = \frac{TP+TN}{(TP+TN+FN+FP)} \]  
\[ \text{Precision} = \frac{TP}{TP+FP} \]  
\[ F - \text{Score} (F_I) = \frac{(1+\beta^2)(\text{Precision}\times\text{Sensitivity})}{(\beta^2)\times(\text{Precision}+\text{Sensitivity})} \]
Where, $\beta$, a default parameter, is commonly 0.5, 1 or 2; in the present study, $\beta = 1$ is being taken.

The odd ratio analysis checks odds of occurrence of an event due to presence of a particular factor (or exposure as called by Pepe et al., 2005) as compared the odds of occurrence of that event in the absence of the same exposure (Pepe et al., 2005).

$$
\text{Odd Ratio} = \frac{\frac{TP}{FP}}{\frac{FN}{TN}}
$$

(30)

The Kappa statistics is a measurement of agreement between two distinguished sets of classification while catering the randomness in the classification (Arabameri et al., 2020a). The Kappa statistics can be computed with the equation which is as follow:

$$
K = \frac{P_{\text{obs}} - P_{\text{exp}}}{1 - P_{\text{exp}}}
$$

(31)

Where, $P_{\text{obs}}$ is observed agreements = (TP+TN), represents the correctly classified values of inundated and non-inundated pixels.

$P_{\text{exp}}$ is expected agreements = $\{(\text{TP+FN}) \times (\text{TP+FP})\} + \{(\text{FP+TN}) \times (\text{FN+TN})\}$ denotes the proportion of inundated and non-inundated pixels which were expected to show agreement, on the basis of chance (Hoehler, 2000).

The value of K ranges between 0 and 1, where the lower value, towards ‘0’, indicates less agreement and higher values, towards ‘1’, shows higher or near to perfect prediction. The various range of K indicates different agreement i.e., $K \leq 0$ (no agreement); 0.01 - 0.20 (slight agreement); 0.21 - 0.40 (fair agreement); 0.41 - 0.60 (moderate agreement); 0.61 - 0.80 (substantial agreement), and 0.81 - 1.00 near to perfect agreement (Cohen, 1960).
The area under the ROC (AUROC) curve measurement describes the evaluated prediction value ranges from 0.5 to 1.0, inaccurate to highly accurate (Marzban, 2004). In this method, the FPR and TPR are plotted on the x-axis and y-axis, respectively. The AUROC values can be classified to measure the accuracy into following four descending order classes: excellent (0.9–1.0), good (0.8–0.9), fair (0.7 to 0.8), and poor (0.6 to 0.7) (Fressard et al., 2014). The AUROC can be calculated as follows (Chapi et al., 2017b):

\[
AUCROC = \sum TP + \sum TN/P + N
\]

(32)

In order to check the errors in the models, the Root Mean Square Error (RMSE) test has been conducted. The RMSE is a standard statistical metric to assess a model’s performance in various subjects of sciences (Chai and Draxler, 2014). The RMSE can be explained mathematically;

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i-O_i)^2}{n}}
\]

(33)

Where the P represents the predicted cases, O refers as observed values, and n is the total number of cases.

In order to check the classification accuracy of the models, the seed cell area index (SCAI) method has been used. The index calculation can be achieved through the ratio of each classified class and the susceptible seed cell percent values (Süzen and Doyuran, 2004).
Where \( N_{pix}(S_{X_i}) \) is the number of pixels with flood occurrence cases within class \( i \) of factor variable \( X \), \( N_{pix}(X_j) \) refers the number of pixels within the factor variable \( X_j \), \( m \) indicates the number of classes in the parameter variable \( X \), and \( n \) represents the number of factors in the study area.

If the low SCAI value fall low for the ‘high’ and ‘very high’ susceptibility classes and high SCAI values represent the ‘low’ and ‘very low’ flood susceptible classes, for a model results, then the model output classification is said to be for that model (Arabameri et al., 2020b).

4. Results and analysis

4.1. Independent analysis of the conditioning factors

The result of MCA for all 12 flood CgFs is presented in table 3. The analysis shows that the TOL values of all 12 CgFs factors are less than 1.0 and the VIF for the same CgFs has been found to be less than 1.9, which suggests that there is no collinearity issue among all 12 CgFs. Therefore, all the factors were further utilized for FSM for prediction.

4.2 Relative weights of conditioning factors

The RF model-based importance of all flood CgFs with their corresponding ranks are presented in table 5. The higher value of a conditioning factor shows their higher importance in FSM (Arabameri et al., 2019f). The results revealed that the highest
predictability in the list of all twelve CgFs, for flood is secured by the geomorphology (0.130). The other major predictors which follow geomorphology, in order of their diminishing importance, are: distance to streams (0.121), curvature (0.103), slope (0.099), stream density (0.094), and LULC (0.082). The other CgFs who play moderate roles in predictability are TWI (0.072), distance to road (0.071), slope aspect (0.070), and altitude (0.070). Two least important predictors are soil (0.050), & rainfall (0.038).

The retrieved values (significance score) and ranking by RF model for all CgFs have been further utilized in prediction of flood using all four models.

4.3 Results of models

The results of all four model-based flood susceptibility models categorized in 5 flood potential classes: very low, low, moderate, high, & very high are displayed in figure 4 (ANFIS), figure 5 (ANFIS-GA), figure 6 (ANFIS-DE), and figure 7 (ANFIS-PSO). For the purpose of presenting the clear picturisation of flood susceptibility near confluence zone, two highly flood susceptible windows have been zoomed-in in all the four maps. All the model results are classified using natural break (NB) method and the class-wise distribution statistics of all the classified map results using NB method is also displayed in the figure 8. In order to have more deeper insights of class-wise areal distribution for each model, the percentage of high & very high classes obtained through different classification methods/schemes viz. Quantile (QNTL), NB, Geometric Interval (GI), & Equal Interval (EI) have also been portrayed in figure 9.

The FSM-ANFIS derived flood susceptibility prediction index (FSPI) values range from 0 to 0.077 which has been divided into five categories using NB method. The
very low (0-0.077) class occupies about 32% of area while ‘low’ class (0.078 to 0.24) covers around 40.58% of area in MGP region. The ‘moderate class’ FSPI ranges from 0.25 to 0.43 and represents 10.40% of area. The least area coverage (about 4.8%) representor is the ‘high class’ (0.44-0.65). And, the ‘very-high class’ (FSPI range: 0.66 -0.779) falls into 12.27% of total region.

The FSM-ANFIS-GA values have also been segmented (using NB) into five classes. The ‘very low-class’ values ranging from 0-0.21, are distributed mainly in higher lands of the basin, occupying 22.49% area. The second category, ‘low class’ values ranging between 0.22 and 0.34, cover 26.15% of the total classified region. About 17.53% of the area is covered by ‘moderate class’ (FSPI range: 0.35-0.49) that mostly occupies the outer parts of the river basin. The fourth class in the list is the high potential range between 0.5 and 0.66 covering 14.53% while the ‘very high’ (FSPI range: 0.67 to 0.99) flood potential category for ANFIS-GA occupies the second highest share, 18.90%, among all incorporated four models.

The values for FSM-ANFIS-DE model-based map range between 0.02 and 0.99; and has been partitioned into five groups using NB. The very low (0.02-0.24) flood potential has spread over 25.06% of the total area in MGP. The highest area of 25.41% has covered by the second category ranging between 0.25 and 0.38. The moderate values range from 0.39 and 0.54 and occupy 22.73% of the total area. The maximum share of 17.29% of the total area of high flood potential values (0.55-0.71) has been predicted by ANFIS-DE among all models. About 9.5% of total area has been claimed by very high (0.72-0.99) flood potential class.

The values of FSM-ANFIS-PSO are included in the interval between 0.06 and 0.94 partitioned into five classes using NB method. The first category, very low values
range between 0.06-0.23 and are distributed mainly in older floodplain of MGP covering 22.82% of the total area. The highest share, 25.48% of the total area, has been covered by the low category ranging from 0.24 to 0.35, expanding mostly in the central part of MGP along with Ganga River. The third category, moderate class, ranging between 0.36 and 0.50, occupies 16.67% of the total area. About 14% of area has been predicted under the ‘high category’ (0.51-0.66) by the ANFIS-PSO model. The ‘very high’ class values ranging between 0.67 and 0.94, secured the top position in terms of predicting the flood potential in MGP with 20.77% of the total area.

The results show that the maximum share of very high class having highest probability of flood occurrences, is predicted by ANFIS-PSO (20.77%) followed by ANFIS-GA (18.90%), ANFIS (12.27%), and ANFIS-DE (9.52%). If examined from another scenario, after merging both the high & very high categories, the maximum share of flooded areas has been secured by ANFIS-PSO (35.03%), ANFIS-GA (33.42%), ANFIS-DE (26.81%), and ANFIS (17.08%). Hence, it can be observed that the ANFIS-PSO has secured the prime position in terms of predicting most high & very high categories of flood occurrences.

In order to compare the results of classifications, another analysis has been done wherein the class-wise distributions of all four models using QNTL, NB, GI, and EI classification schemes, it is found that the ANFIS-PSO & ANFIS-GA have shown better performance than other two models in terms of predictability of the ‘high & very high classes’ of flood susceptible lands. The ANFIS model has found to be the least performant in terms of sharing the high class using NB (4.81%) and EI (5.26%) classification schemes whereas, ANFIS-DE model has been found to be at the end of the list for sharing the very high class using EI (5.78%) and NB (9.52%).
4.4 Evaluation of the FSM

4.4.1 Cut-off-independent evaluation metric: AUCROC

The AUC evaluates the overall performance of a model (Jaafari et al., 2019; Khosravi et al., 2019; Liu and Li, 2005; Pham et al., 2019). In the present study, a promising and acceptable level of AUC was found in the case of all four models in both with respect to training (success rate) and validation (prediction rate). The AUC\textsubscript{SR} (success rate AUC) ranges between 0.807 and 0.922 whereas the AUC\textsubscript{PR} (prediction rate AUC) was found between 0.768 and 0.924 for ANFIS model and its three ensembles. The highest position for success (AUC = 0.922) and prediction (AUC = 0.924) was secured by ANFIS-GA. The ANFIS-PSO positioned at second in the list with AUC=0.915 (success) and AUC = 0.921 (prediction). Among the other remaining two models, ANFIS (AUC\textsubscript{SR} = 0.807; AUC\textsubscript{PR}=0.768) & ANFIS-DE (AUC\textsubscript{SR} = 0.901; AUC\textsubscript{PR}=0.919), the latter had been found to be more accurate than the former in terms of training & testing accuracy. The success rate curve (SRC) and prediction rate curves (PRC) are shown in Figures 10A and 10 B, respectively.

4.4.2 Cut-off-dependent evaluation matrices

In order to calculate the cut-off-dependent matrices, the collected True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) samples were utilized. The complete results are provided in table 6.

4.4.2.1 Accuracy (Ac)

The first in the list of dependent evaluation matrices is accuracy, which defines that how accurately the model has performed during training and validation of the data. This metric’s values range between 0 and 1; the higher value indicates more
accuracy and the vice versa. The maximum accuracy has been observed for ANFIS-GA for training (0.886) and validation (0.883), followed by ANFIS-PSO with accuracy levels of 0.871 for training and 0.875 for validation whereas the lowest accuracy was seen in case of ANFIS model with accuracy values of 0.855 for training and 0.853 for validation.

4.4.2.2 Sensitivity (Sn)

The Sn is calculated by dividing the number of true positive predictions by the total positive values (sum of true positive & false negative values). The highest sensitivity was recorded during training of the model for ANFIS-GA (0.886) followed by ANFIS-PSO, ANFIS-DE (0.867), and ANFIS (0.863). During validation process, the ANFIS-GA & ANFIS-PSO shared the first position with sensitivity figure of 0.867. The second position was secured by ANFIS-DE (0.858) and in the last position; the ANFIS (0.847) has been placed.

4.4.2.3 Specificity (Sp)

The Sp is calculated by dividing the total number of true negative predictions by the total negative counts (sum of false positive & true negative values). The specificity also refers as precision of the model. The prime position during training of model was secured by ANFIS-GA (0.887) following with ANFIS-PSO (0.876), ANFIS-DE (0.869) and ANFIS (0.847). Again, during validation, the ANFIS-GA (0.900) outhustled the other three models; ANFIS-PSO (0.883), ANFIS-DE (0.867) and ANFIS (0.860).

4.4.2.4 F-Score
In order to provide more insights about performance comparison using dependent evaluation metrics for each model, the F-score is also computed in the study; it analyses the harmonic mean score through the calculated precision and sensitivity of the data (Rahmati et al., 2019). The ANFIS-GA has received the highest F-Score during training (0.886) and validation (0.881). The ANFIS-PSO has also performed better in comparison to other two models and had scored very well during training (0.871) and validation (0.874). The third position was secured by ANFIS-DE during training (0.868) and validation (0.861). The least performant was ANFIS which was found positioned at the last in the list during training (0.856) and validation (0.852) procedures.

4.4.2.5 Cohen’s Kappa (K)

By analysing the K-metric of all the models during training and validation, it was found that the consistent performing model was ANFIS-GA. During training, the top rank was secured by ANFIS-GA (0.773); with the following successors: ANFIS-PSO (0.743), ANFIS-DE (0.736) and ANFIS (0.710). The ANFIS-GA secured the first place with K = 0.767 during validation followed by ANFIS-PSO (0.750), ANFIS-DE (0.723), and ANFIS (0.707).

4.4.2.6 RMSE

The RMSE has been recorded to be <0.42 (thereby meaning low error) for all the models during training and validation. The lowest error was recorded, during training, by ANFIS-GA (0.398), with successors: ANFIS-PSO (0.402), ANFIS-DE (0.410), and ANFIS (0.413). The same ranking was also seen, during validation part, for all models. The most successful model ANFIS-GA has recorded lowest error,
RMSE=0.375, following by ANFIS-PSO (0.379), ANFIS-DE (0.387), and ANFIS (0.390).

Apart from above important matrices results, the supplementary parameters i.e. false positive rate, false discovery rate, false negative rate, and Odd-Ratio are also provided in the table 5.

4.4.3 SCAI

Another validation assessment has been performed through SCAI method to check the classification accuracy of produced maps obtained using all the four models. The classification accuracy results are displayed in figure 11. The SCAI calculates the ratio between the class share and the observed flood share for the particular class. And after the analysis of the SCAI values are found low for low susceptible classes (Very low to Moderate classes), it indicates that the classification accuracy is poor for that particular map; also, the high SCAI values for high susceptible classes (High & Very High classes) refers the same low classification accuracy as well. But the results found in this study show that the high SCAI values have been observed for low susceptible classes as well as low SCAI values for high susceptible classes and it proves that the produced maps of all the models are of good to excellent classification accuracy category. The maximum classification accuracy was observed in case of ANFIS-PSO & ANFIS-GA models. The ANFIS based modelled FSPI map shows less accuracy in comparison to the other three models. In order to have better comprehension of the model accuracies, the frequency ratio (FR) based diagram (figure 11A) is also provided along with SCAI diagram (figure 11B).

5. Discussion
5.1 Assessment of variable importance

The assessment of CgFs importance is relevant for planners and decision makers for better utilization of resources and maximum productivity using limited resources allocation (Testa et al., 2016). The derivation of importance of CgFs for spatial modelling using bivariate models computation at the time of analysing their class-wise importance or weightages assignment with the aid of frequency ratio, evidential belief function, Shannon’s entropy, Index of entropy, etc. (Arabameri et al., 2019a; Arora et al., 2019). But, application of ML based models requires a pre-analysed weightage or importance assignment to all the CgFs before employment of models for spatial modelling. There are a number of studies available where the relative importance of variables has been computed before spatial modelling using different methods i.e. Analytic Hierarchy Process (AHP) (Pradhan, 2017), Information-Gain method (Costache et al., 2020b), random forest model (Arabameri et al., 2019f), Stepwise Weight Assessment Ratio Analysis (SWARA) (Bui et al., 2018b; Chen et al., 2019), etc. And, there are numerous approaches being continuously used to analyse the relative importance of CgFs for spatial modelling. The performance of different techniques have also been assessed on the same set of CgFs to select the best ones from among list but the importance of CgFs is also affected and determined by the type of topography and data utilized for derivation of those CgFs. Therefore, the performance of models is also compromised due to various factors involved in preparation of CgFs. Hence, the selection of a method for assessing relative importance is not strict or limited to any isolated model. In present work, the random forest model has been used for evaluation of relative importance of each conditioning factor. The validation of the model performance was performed using confusion matrix table (table 4). An amicably acceptable level of accuracy and
precision (~0.70 for both parameters) was recorded by the RF model using the
confusion matrix table.

Here, from the list, we can observe that the geomorphology has outperformed all the
other CgFs in terms of contributor to the predictability of flood susceptibility. Since,
the MGP is predominantly a fluvial environment and the geomorphology dataset
carries the most valuable information related to the fluvial form-process relationship
represented in form of geomorphological units e.g. active & old floodplain,
paleochannel, channel bar, bars, meanders, etc. as well as provide a layout of
historical flood events, the model results also verify the predictability importance of
geomorphology in its result. The second important factor, distance to stream, in the
list as a predictor of flood can be explained as the streams or rivers are prime source
of flood in the fluvial environment, especially in plains like the MGP, bringing havoc
at recurring interval. At the same time, the curvature and slope are also found as
important factors for flood predictability due to almost plain surface topography in the
study area. Here, it may be noted that the rainfall, one of the major flood-controlling
factors, secured the last position in the analysis. It can be explained as the least
factor for flood in MGP due to the location of MGP. The MGP lies in the lower Ganga
basin and here three major rivers- Ganga, Ghaghara & Son form a confluence zone,
almost in the middle of the study area, means the flood water during monsoon period
have been brought by upstream channels including these three major rivers. In fact,
when rainfall occurs in upstream channels, the heavy streams bring flooded water in
downstream channel stretches of the Ganga river basin. Therefore, in MGP the
rainfall couldn’t come up as a chief controlling factor for flood.

5.2 The selection and impact of conditioning factors in FSM
The success of a prediction model is dependent, directly or indirectly, on the selection procedure and type of the conditioning factors. It also determines prediction quality of involved models (Arora et al., 2019; Lee et al., 2017). It can be noted that there is no universal guideline available for selection of conditioning factors as well as their number and class divisions (Khosravi et al., 2018). But the selection of flood controlling CgFs can be affected by the type of topography and climatic conditions. It has been elaborated in the previous study of FSM using bivariate models that the geomorphology has performed better in comparison to geology (lithology) as a contributor to floods in the fluvial floodplain topographic setting such as the MGP (Arora et al., 2019); therefore, in present study the more refined version of geomorphology dataset has been selected. Including geomorphology, total twelve CgFs have been selected in this study keeping in mind the existing fluvial setup and hydro-climatological conditions of the MGP. The preferability of geomorphology over geology is also justified by the variable importance analysis using RF model. Wherein, the geomorphology as a contributing factor is listed on top with the maximum weightage as opposed to other controlling factors for flood. The other impactful CgFs have been noted, after factor importance analysis results which are stream density, distance to stream, TWI, Slope, and curvature.

5.3 The model performance evaluation and comparison

All the heuristic models, ANFIS & their ensembles, performed very well and the assessment of their performance were also verified by the AUCROC method including other dependent metrices. The AUC for all four models ranges between 0.768 and 0.924 during training and validation parts. The most promising model, ANFIS-GA_{SR} (AUC=0.922) & ANFIS-GA_{PR} (AUC=0.924), has secured the top position in comparison to remaining models. There has been no potential study
conducted in MGP or in any other similar topographic settings in India to date for FSM by using the ANFIS & ensembles models, therefore, in comparison with other similar kind of work it has been observed that the performance of results are found better to the study conducted by Hong et al. (2018) using ANFIS-DE ($AUC_{SR} = 0.8523, AUC_{PR} = 0.8686$) & ANFIS-GA ($AUC_{SR} = 0.8488, AUC_{PR} = 0.8743$) models in Hengfeng County, China (Hong et al., 2018a). The varied difference in both results using same models may occur due to different topographic configuration and selection of CgFs. In comparison to another study conducted by M. Ahmadlou et al. (2019) for Iran using ANSIF & ensembles using biogeography-based optimization (BBO) and BAT algorithm (BA) (ANFIS-BBO & ANFIS-BAT) (Ahmadlou et al., 2019), the performance of used models in present study have found much better during training & validation. In earlier work, the maximum AUC was recorded for both ANFIS-BBO & ANFIS-BAT ($AUC=0.77$) during training and $AUC=0.70$ was recorded during testing for both models. Whereas, in present work, even the lowest performing ANFIS model has performed better than both models.

In the study conducted by Hong et al. (2018), the altitude range recorded at some places are ~1340 meter which indicates that the topography of the study area was not plain topography compared to the MGP topographic setting where the maximum altitude is recorded ~100 meter only and the elevation range is ~83m. Another important point observed between both studies is that the present study used the geomorphology as a flood controlling factor keeping in mind fluvial topographic setting in the present MGP area whereas, Hong et al. did not.

Apart from these major geo-environmental factors’ selection, other reasons for variations in model performance maybe counted in the quality of preparation of flood inventory, the number of flood and non-flood point generation and utilization of
training & testing points as samples, data processing, and preparation of CgFs using satellite images; where the calibration, resolution (spatial & temporal), & size of input data (satellite images or in-situ data).

Having acquired all the performance facets of models employed in this study, it will be helpful to the hazard managers, both national and local level managers, to have an idea about which areas are most susceptible to flooding and which ones are least. Also, the sequence of contributing factors used for flood occurrence potential prediction or flood susceptibility prediction, the authorities will be able to priorities their efforts to apply nature based solutions (NBS) like creating natural embankments, or making tree barriers, etc. or other NBS as per the suitability of the measures’ to help curb the detrimental effects of the factors. Gómez Martín et al., (2020) have suggested two classifications of NBS: 1) horizontal NBS framework; and 2) vertical NBS framework. Horizontal NBS framework advocate for levels of human intervention in the solution measures and based on this, three types of horizontal NSB are suggested: a) NBS type 1; b) NBS type 2; and c) NBS type 3; NBS type 1 involves low human intervention and type 3 involves higher levels of modifications in the ecosystem. Green, blue, or grey infrastructural measures of NBS are suggested for city based flood inundation related management practices (Commission, 2015; Eggermont et al., 2015, 2015; Faivre et al., 2017). Geomorphology being the most important factor contributing to the flood susceptibility prediction in this low altitudinal subtropical monsoonal topoclimatic regions, the NBS of simplest type will be to identify geomorphic units which are flooded regularly and which are most inhabited or occupied for economic activities, NBS types 1, 2, or 3, should be applied for sustainable development of the area.

6. Conclusions
The Central Gangetic floodplains being one of the worst flood affected regions in the world still lag proper natural hazard actionable policies crafted to tackle the losses due to this menace. In order to execute the effectively actionable decision-making process in this critically flood affected region, aim should be reducing losses related to hydraulic projects for sustainable development, developing sites for industrial hub demarcation and settlement areas, the updated, accurate, verifiable flood prone areas, their susceptibility, etc. are of prime significance. The present work has been carried out with an impetus to fulfil the data lag in this frequently flooded part of the MGP region located in the Ganga River Basin, India, during monsoon period (June to October), improvements in the model performances warrants an improved quality of model input parameters. In our present work, the inclusion of upgraded geomorphology dataset as a higher quality conditioning factor has helped in improving the accuracy of the model ensembles. The RF model analysis has revealed the highest relative importance of the geomorphology among all CgFs. Comparison of results with the previous study conducting the FSM in the same part of the MGP applying FR & Shannon’s entropy (SE)(Arora et al., 2019) signifies the fact that there is achieved an improvement is achieved in the performance of the advanced and novel ensemble models, not previously used by any workers for flood susceptibility zonation in any part of India. The present work used the ANFIS & novel ensembles and recorded better accuracy than models used in earlier study in the same topoclimatic environmental setting. Among all the four used models in present work, the ANFIS-GA has come up to be the best performant after assessing all models’ training & testing performances by independent (AUROC) & multiple dependent metrics methods i.e. accuracy, prediction, F-Score, Cohen’s Kappa, etc. The results and comparison also show that the selection & optimization of CgFs may
also improve the accuracy of models for prediction. It is also noted that the ensembles have performed better than the isolated individual ANFIS model. The efficacy and accuracy of all ensembles have also been noted better during training and testing stages. Also, the coverages of high susceptible lands for flood resonated better with flood samples for the ensemble-based produced maps than ANFIS based map.

Like other studies aiming to achieve the model universality, this study too has its own constraints and limitations related to DEM data quality, model configuration related limitations, algorithms’ assumptions, etc. but more and better advancement on these frontiers of flood susceptibility zonation modelling is still under progress and more such studies with better field related data will be needed to help improve the precision and accuracy of such modelling exercises aimed at sustainable development of such flood affected areas.

Studies like this can be viewed as beneficial in two time frames: 1) present day; and 2) future. The future flood menace presenting challenges to the natural hazard management authorities in the present timeframe will help in gaining information related to what factors are more important contributors in flood occurrence potential prediction and thus, help to focus their attention to take nature based solution measures to control the flood waters reaching such places. And secondly, the flood susceptibility zonation maps will lead to overall fund allocation for different zones depending upon the potential damages probable in those zones. As far as future flood scenarios are concerned, since we are heading towards an era of ‘space based monitoring, analysis, assessment and evaluation’ of all natural (both biotic and abiotic) and manmade phenomena, the intensive testing (Abdelghafar et al., 2020; Alsdorf et al., 2003; DeVries et al., 2020; Green et al., 2020; Hassanien et al., 2020;
Kummerow et al., 2020; Malinowska et al., 2020; Salisbury et al., 2013; Spyropoulos et al., 2020; Sterckx et al., 2020; Tang et al., 2010; Thies and Bendix, 2011; Tonetti et al., 2020), and different machine learning algorithms are being developed and already existing ones are being tested and validated considering all possible scenarios including different types of topoclimatic settings. In flood hazard management science, the modelling community is engaged in developing, testing and validating models for different environmental settings. But, till date, this dimension of model testing as per variability of topoclimatic and hydrometeorological differences has recently been mulled and promulgated by Arora et al. 2019. Arora and his team have attempted developing, testing and validating all types of susceptibility models, starting from simplest ‘frequency ratio (FR)’ and ‘Shannon Entropy (SE)’ models to more complex ones like for low altitudinal range floodplain environment. There is a need to test all existing types of models e.g. other machine learning standalone and ensembles of MLP (multiplayer perceptron), LR (logistic regression), CART (classification and regression tree), SVM (support vector machine), etc. with other categories of models like FR, EBF (evidential belief function), etc. After critical comparison of all types of models’ performances in different topoclimatic and hydrometeorological settings, the best performing models may be suggested to be set in automated future satellite missions that will able to provide Near Real Time Flood Monitoring and Susceptibility Prediction viz. SWOT satellite mission (Morrow et al., 2019) that provides topographic and surface water information at very fine temporal and spatial resolutions for the world oceans.

References


Computational Intelligence Models for Improvement Gully Erosion Assessment.

Remote Sens. 12, 140. doi:10.3390/rs12010140


Arabameri, A., Yamani, M., Pradhan, B., Melesse, A., Shirani, K., Tien Bui, D., 2019f. Novel ensembles of COPRAS Multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of...


Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250. doi:10.5194/gmd-7-1247-2014

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., Khosravi, K., 2017a. CO.


Nearest Neighbor Classifier. doi:10.3390/rs12020266


Thies, B., Bendix, J., 2011. Satellite based remote sensing of weather and climate:


List of figures

**Figure 1:** (A) Location map of Middle Ganga Plain (MGP); (B) Mean monthly discharge during monsoon period (Jun to October) at Gandhighat station for 2008 year.
Figure 2: Flowchart of methodological steps followed in this work
Figure 3: Flood conditioning factors. From A to L the maps indicate: Altitude; Curvature; Distance to Road; Distance to Stream; Geomorphology; Land Use Land Cover; Rainfall; Slope Aspect; Slope; Soil Type; Stream Density; and Topographical Wetness Index (TWI).
Figure 4: Flood susceptibility mapping using ensemble of adaptive neuro-fuzzy inference system (ANFIS) model

Figure 5: Flood susceptibility mapping using ensemble of adaptive neuro-fuzzy inference system (ANFIS) model with genetic algorithm (GA)
Figure 6: Flood susceptibility mapping using ensemble of adaptive neuro-fuzzy inference system (ANFIS) model with a) differential evolutionary (DE)

Figure 7: Flood susceptibility mapping using ensemble of adaptive neuro-fuzzy inference system (ANFIS) model with particle swarm optimization (PSO)
Figure 8: Percentage of each susceptibility classes classified using natural break in all models’ results

Figure 9: Percentage of ‘High’ & ‘Very High’ susceptibility classes of flood susceptibility potential mapping as per all the models suing different classification methods (QNTL=Quantile; NB=Natural Break; GI=Geometric Interval; EI=Equal Interval)
Figure 10: Area under the receiver operating characteristic (ROC) curves for all the models; Section (A) is with training data (success rate curve) and section (B) has been constructed with the validation data (prediction rate curve).

Figure 11: Validation of results using (A) frequency ratio (FR) and (B) seed cell area index (SCAI)

List of tables
Table 1: Details of satellite data used in the study

<table>
<thead>
<tr>
<th>SN</th>
<th>Event time</th>
<th>Satellite (Spatial Resolution)</th>
<th>Acquisition date</th>
<th>Spatial reference (Projected)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Preflood</td>
<td>Landsat 5 TM (30-meter)</td>
<td>28-05-2008</td>
<td>Projection: UTM</td>
</tr>
<tr>
<td>2</td>
<td>During Flood</td>
<td>SRTM v4 DEM (1-arc or 30-meter)</td>
<td>01-09-2008</td>
<td>Datum: WGS84</td>
</tr>
<tr>
<td>3</td>
<td>Post Flood</td>
<td></td>
<td>19-10-2008</td>
<td>Spheroid: WGS84</td>
</tr>
<tr>
<td>4</td>
<td>----</td>
<td></td>
<td>11-22 Feb 2000</td>
<td>Zone: 44N</td>
</tr>
</tbody>
</table>

Table 2: Geomorphological units in study area mapped and used in the modelling as one of the input conditioning factors

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Geomorphic Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>FluOri - Active Flood plain</td>
<td>Junk et al. (1989) have presented, following Bhowmik &amp; Stall (1979), active floodplain in technical terms and stated that the areas flooded by 100-year flood can be demarcated to be under the active floodplain zone. Active floodplain an overflown surface, usually, on either side of a stream which is periodically flooded and experiences both erosional and depositional processes building and destroying the surfaces but net result is net surface growth through the accretion of the depositional material (Williams, 1978). It should be noted that some river reaches do not have these active floodplain surfaces. These geomorphological units have high flood susceptibility as compared to older floodplain surfaces also termed as terraces.</td>
</tr>
<tr>
<td>2.</td>
<td>Meander scar</td>
<td>Meander scar is a crescent-shaped fossilized meandering channel segment currently covered with vegetation but with conspicuously visible channel geometry in the satellite imagery. Such geomorphic features represent low lying areas exposed to more susceptibility to</td>
</tr>
</tbody>
</table>
flooding (Harris, 1987).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td><strong>Braid Bar</strong></td>
</tr>
<tr>
<td>4.</td>
<td><strong>Lateral Bar/lateral channel bar</strong></td>
</tr>
<tr>
<td>5.</td>
<td><strong>Marsh</strong></td>
</tr>
<tr>
<td>6.</td>
<td><strong>Channel Island</strong></td>
</tr>
<tr>
<td>7.</td>
<td><strong>Paleochannel</strong></td>
</tr>
<tr>
<td>8.</td>
<td><strong>WatBod - Pond</strong></td>
</tr>
<tr>
<td>9.</td>
<td><strong>FluOri - Older Flood plain</strong></td>
</tr>
<tr>
<td>10.</td>
<td><strong>Abandoned Channel</strong></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>11.</td>
<td>Point Bar</td>
</tr>
<tr>
<td>12.</td>
<td>FluOri - Older Alluvial Plain</td>
</tr>
<tr>
<td>13.</td>
<td>Channel Bar</td>
</tr>
<tr>
<td>14.</td>
<td>WatBod - River</td>
</tr>
<tr>
<td>15.</td>
<td>Cut-off Meander</td>
</tr>
<tr>
<td>16.</td>
<td>Backswamp</td>
</tr>
<tr>
<td>17.</td>
<td>Valley Fill</td>
</tr>
<tr>
<td>18.</td>
<td>Oxbow Lake</td>
</tr>
<tr>
<td>19.</td>
<td>Natural Levee</td>
</tr>
<tr>
<td>20.</td>
<td>FluOri - Younger Alluvial plain</td>
</tr>
</tbody>
</table>
Repeatedly flooded zone, on both sides channel banks, adjacent to the current active channel which get annual alluvial deposits marks the younger alluvial floodplain.

| 21. | WatBod - Others | Other smaller waterbodies other than river and pond. |

Note: The data source is the Geological Survey of India, Government of India official portal (http://bhukosh.gsi.gov.in/), but the contents of this map has been prepared and updated by author using different sources.
Table 3: Multicollinearity Analysis of the flood conditioning factors

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Conditioning Factors</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td>1.</td>
<td>Altitude (Al)</td>
<td>0.538</td>
</tr>
<tr>
<td>2.</td>
<td>Curvature (Cr)</td>
<td>0.924</td>
</tr>
<tr>
<td>3.</td>
<td>Distance to Roads (D2R)</td>
<td>0.926</td>
</tr>
<tr>
<td>4.</td>
<td>Distance to Streams (D2S)</td>
<td>0.776</td>
</tr>
<tr>
<td>5.</td>
<td>Geomorphology (G)</td>
<td>0.913</td>
</tr>
<tr>
<td>6.</td>
<td>Land Use Land Cover (LULC)</td>
<td>0.707</td>
</tr>
<tr>
<td>7.</td>
<td>Rainfall (Rf)</td>
<td>0.906</td>
</tr>
<tr>
<td>8.</td>
<td>Slope (Sl)</td>
<td>0.749</td>
</tr>
<tr>
<td>9.</td>
<td>Slope Aspect (SA)</td>
<td>0.874</td>
</tr>
<tr>
<td>10.</td>
<td>Soil (S)</td>
<td>0.784</td>
</tr>
<tr>
<td>11.</td>
<td>Stream Density (SD)</td>
<td>0.733</td>
</tr>
<tr>
<td>12.</td>
<td>Topographic Wetness Index (TWI)</td>
<td>0.623</td>
</tr>
</tbody>
</table>

Table 4: Confusion matrix from the RF model (1 = flood, 0 = non-flood)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>325</td>
<td>167</td>
</tr>
<tr>
<td>1</td>
<td>130</td>
<td>357</td>
</tr>
</tbody>
</table>

Table 5: Relative influence of effective conditioning factors in the RF model

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Factor</th>
<th>Relative weight</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Altitude (Al)</td>
<td>0.07</td>
<td>10</td>
</tr>
<tr>
<td>2.</td>
<td>Curvature (Cr)</td>
<td>0.10</td>
<td>3</td>
</tr>
<tr>
<td>3.</td>
<td>Distance to Roads (D2R)</td>
<td>0.07</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Distance to Streams (D2S)</td>
<td>0.12</td>
<td>2</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------</td>
<td>------</td>
<td>---</td>
</tr>
<tr>
<td>5.</td>
<td>Geomorphology (G)</td>
<td>0.13</td>
<td>1</td>
</tr>
<tr>
<td>6.</td>
<td>Land Use Land Cover (LULC)</td>
<td>0.08</td>
<td>6</td>
</tr>
<tr>
<td>7.</td>
<td>Rainfall (Rf)</td>
<td>0.04</td>
<td>12</td>
</tr>
<tr>
<td>8.</td>
<td>Slope (Sl)</td>
<td>0.10</td>
<td>4</td>
</tr>
<tr>
<td>9.</td>
<td>Slope Aspect (SA)</td>
<td>0.07</td>
<td>9</td>
</tr>
<tr>
<td>10.</td>
<td>Soil (S)</td>
<td>0.05</td>
<td>11</td>
</tr>
<tr>
<td>11.</td>
<td>Stream Density (SD)</td>
<td>0.09</td>
<td>5</td>
</tr>
<tr>
<td>12.</td>
<td>Topographic Wetness Index (TWI)</td>
<td>0.07</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 6: Validation of results by model evaluation metrics

<table>
<thead>
<tr>
<th>Data</th>
<th>Criteria</th>
<th>ANFIS-PSO</th>
<th>ANFIS</th>
<th>ANFIS-DE</th>
<th>ANFIS-GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Negative (TN)</td>
<td>613</td>
<td>593</td>
<td>608</td>
<td>621</td>
</tr>
<tr>
<td></td>
<td>False Positive (FP)</td>
<td>87</td>
<td>107</td>
<td>92</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>False Negative (FN)</td>
<td>93</td>
<td>96</td>
<td>93</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>True Positive (FP)</td>
<td>607</td>
<td>604</td>
<td>607</td>
<td>620</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>0.871</td>
<td>0.855</td>
<td>0.868</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.876</td>
<td>0.847</td>
<td>0.869</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>0.867</td>
<td>0.863</td>
<td>0.867</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>0.876</td>
<td>0.847</td>
<td>0.869</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>False Positive Rate</td>
<td>0.1243</td>
<td>0.1505</td>
<td>0.1316</td>
<td>0.1130</td>
</tr>
<tr>
<td></td>
<td>False Discovery Rate</td>
<td>0.1254</td>
<td>0.1529</td>
<td>0.1314</td>
<td>0.1129</td>
</tr>
<tr>
<td></td>
<td>False Negative Rate</td>
<td>0.1329</td>
<td>0.1393</td>
<td>0.1327</td>
<td>0.1141</td>
</tr>
<tr>
<td></td>
<td>F-Score</td>
<td>0.871</td>
<td>0.856</td>
<td>0.868</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>Odd ratio</td>
<td>45.988</td>
<td>34.869</td>
<td>43.134</td>
<td>60.921</td>
</tr>
<tr>
<td></td>
<td>Cohens kappa</td>
<td>0.743</td>
<td>0.710</td>
<td>0.736</td>
<td>0.773</td>
</tr>
<tr>
<td></td>
<td>AUCROC</td>
<td>0.915</td>
<td>0.807</td>
<td>0.901</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.402</td>
<td>0.413</td>
<td>0.410</td>
<td>0.398</td>
</tr>
<tr>
<td>Training data</td>
<td>True Negative (TN)</td>
<td>265</td>
<td>258</td>
<td>260</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>False Positive (FP)</td>
<td>35</td>
<td>42</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>False Negative (FN)</td>
<td>40</td>
<td>46</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>True Positive (FP)</td>
<td>260</td>
<td>254</td>
<td>257</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>0.875</td>
<td>0.853</td>
<td>0.862</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.883</td>
<td>0.860</td>
<td>0.867</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>0.867</td>
<td>0.847</td>
<td>0.857</td>
<td>0.867</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
<td>0.883</td>
<td>0.860</td>
<td>0.867</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>False Positive Rate</td>
<td>0.1186</td>
<td>0.1419</td>
<td>0.1347</td>
<td>0.1034</td>
</tr>
<tr>
<td></td>
<td>False Discovery Rate</td>
<td>0.1167</td>
<td>0.1400</td>
<td>0.1333</td>
<td>0.1000</td>
</tr>
<tr>
<td></td>
<td>False Negative Rate</td>
<td>0.1311</td>
<td>0.1513</td>
<td>0.1419</td>
<td>0.1290</td>
</tr>
<tr>
<td></td>
<td>F-Score</td>
<td>0.874</td>
<td>0.852</td>
<td>0.861</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>Odd ratio</td>
<td>49.214</td>
<td>33.919</td>
<td>38.849</td>
<td>58.500</td>
</tr>
<tr>
<td></td>
<td>Cohens kappa</td>
<td>0.750</td>
<td>0.707</td>
<td>0.723</td>
<td>0.767</td>
</tr>
<tr>
<td></td>
<td>AUCROC</td>
<td>0.921</td>
<td>0.768</td>
<td>0.919</td>
<td>0.924</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.379</td>
<td>0.390</td>
<td>0.387</td>
<td>0.375</td>
</tr>
</tbody>
</table>

| Validation data | True Negative (TN) | 265 | 258 | 260 | 270 |
|                | False Positive (FP) | 35  | 42  | 40  | 30  |
|                | False Negative (FN) | 40  | 46  | 43  | 40  |
|                | True Positive (FP)  | 260 | 254 | 257 | 260 |
|                | Accuracy            | 0.875 | 0.853 | 0.862 | 0.883 |
|                | Precision           | 0.883 | 0.860 | 0.867 | 0.900 |
|                | Sensitivity         | 0.867 | 0.847 | 0.857 | 0.867 |
|                | Specificity         | 0.883 | 0.860 | 0.867 | 0.900 |
|                | False Positive Rate | 0.1186 | 0.1419 | 0.1347 | 0.1034 |
|                | False Discovery Rate| 0.1167 | 0.1400 | 0.1333 | 0.1000 |
|                | False Negative Rate | 0.1311 | 0.1513 | 0.1419 | 0.1290 |
|                | F-Score             | 0.874 | 0.852 | 0.861 | 0.881 |
|                | Odd ratio           | 49.214 | 33.919 | 38.849 | 58.500 |
|                | Cohens kappa        | 0.750 | 0.707 | 0.723 | 0.767 |
|                | AUCROC              | 0.921 | 0.768 | 0.919 | 0.924 |
|                | RMSE                | 0.379 | 0.390 | 0.387 | 0.375 |