How Do Multiple Kernel Functions in Machine Learning Algorithms Improve Precision in Flood Probability Mapping?

Muhammad Aslam Baig
Institute of Mountain Hazards and Environment Chinese Academy of Sciences https://orcid.org/0000-0002-8454-478X

Donghong XIONG
Institute of Mountain Hazards and Environment Chinese Academy of Sciences https://orcid.org/0000-0002-6962-3551

Mahfuzur Rahman (mfz.rahman@iubat.edu)
International University of Business Agriculture and Technology

Md. Monirul Islam
International University of Business Agriculture and Technology

Ahmad Elbeltagi
Mansoura University

Belayneh Yigez
Institute of Mountain Hazards and Environment Chinese Academy of Sciences

Dil Kumar Rai
Institute of Mountain Hazards and Environment Chinese Academy of Sciences

Muhammad Tayab
Northeast Normal University

Ashraf Dewan
Curtin University Bentley Campus: Curtin University

Research Article

Keywords: Hydro-climatic hazards, Machine learning algorithms, Gaussian process regression, Support vector machine, Climate change

Posted Date: August 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-749595/v1

How do multiple kernel functions in machine learning algorithms improve precision in flood probability mapping?

Muhammad Aslam Baig[a,b], Donghong Xiong[a,*], Mahfuzur Rahman[c]*, Md. Monirul Islam[c], Ahmed Elbeltagi[d,e], Belayneh Yigez[a,b], Dil Kumar Rai[a,b], Muhammad Tayab[f], Ashraf Dewan[g]

[a]Key Laboratory for Mountain Hazards and Earth Surface Process, Institute of Mountain Hazards and Environment (IMHE), Chinese Academy of Sciences (CAS), Chengdu 610041, China

[b]University of Chinese Academy of Sciences (UCAS), Beijing 100049, China

[c]Department of Civil Engineering, International University of Business Agriculture and Technology (IUBAT), Dhaka 1230, Bangladesh

[d]Agricultural Engineering Dept., Faculty of Agriculture, Mansoura University, Mansoura, 35516, Egypt

[e]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, 310058, China

[f]Institute of Natural Disaster Research, School of Environment, Northeast Normal University, Changchun 130024, China.

[g]School of Earth and Planetary Sciences, Curtin University, Kent St, Bentley, WA 6102, Australia

*Corresponding author: dhxiong@imde.ac.cn (D.H. Xiong)

Abstract With climate change, hydro-climatic hazards, i.e., floods in the Himalayas regions, are expected to worsen, thus, likely to affect humans and socio-economic growth. Precisely, the Koshi River basin (KRB) is often impacted by flooding over the year. However, studies on estimating and predicting floods still lack in this basin. This study aims at developing flood probability map using machine learning algorithms (MLAs): gaussian process regression (GPR) and support vector machine (SVM) with multiple kernel functions including Pearson VII function kernel (PUK), polynomial, normalized poly kernel, and radial basis kernel function (RBF). Historical flood locations with available topography, hydrogeology, and environmental datasets were further considered to build flood model. Two datasets were carefully chosen to measure the feasibility and robustness of MLAs: training dataset (location of floods between 2010 and 2019) and testing dataset (flood locations of 2020) with thirteen flood influencing factors.

The validation of the MLAs was evaluated using a validation dataset and statistical indices such as the coefficient of determination (r^2: 0.546~0.995), mean absolute error (MAE: 0.009~0.373), root mean square error (RMSE: 0.051~0.466), relative absolute error (RAE: 1.81~88.55%), and root-relative square error (RRSE: 10.19~91.00%). Results showed that the SVM-Pearson VII kernel (PUK) yielded better prediction than other algorithms. The resultant map from SVM-PUK revealed that 27.99% area with low, 39.91% area with medium, 31.00% with high, and 1.10% area with very high probabilities of flooding in the study area. The final flood probability map could add a greatt value to the effort of flood risk mitigation and planning processes in KRB.

Keywords: Hydro-climatic hazards; Machine learning algorithms; Gaussian process regression; Support vector machine; Climate change

1. Introduction

Climate change has led to intense rainfall and floods that significantly influence lives, properties, and socio-economic progress (Tabari, 2020). Of course, flood is one of the most frequent and expensive in terms of human and economic losses among environmental disasters and comprise thirty-one percent of worldwide financial damage in total, caused by environmental disasters (Yalçın, 2002). According to the Sendai Framework for Disaster Risk Reduction (DRR), countries should significantly minimize disaster risk and losses over the next 15 years (Msabi and Makonyo, 2021). This can only be accomplished through relevant research and conflict resolution between scientists and policymakers at all levels.

Flood probability mapping (FPM) is essential for designing, building, and administering appropriate countermeasures. However, owing to complexity of the processes, such as the runoff formation mechanism, climate variation, and the effect of anthropogenic activities (land cover change, population growth, etc.), it is challenging to simulate floods with acceptable precision (Dano et al., 2019). Thus, developing a model of flood probability that is more consistent with the natural habitat and can account for changing meteorological conditions is the focus of basin-scale flood research.

The Koshi River basin's water-related incidents are glacial lake outbursts, flooding, debris flow, drought, and so forth. (Chen et al., 2017). These hazards exhibit high frequencies, large impacts and prolonged durations. The Koshi River basin (KRB) is prone to flooding, and in the last 60 years, Nepal has been hit by ten large-scale floods

triggered by the Koshi river (Dixit, 2009). These floods are the result of interaction between humans and the environment. The frequent rainstorms in the basin contribute significantly to the increasing flood hazards. Besides, miscommunication exists in the operation of hydraulic infrastructure led to flooding. Precisely, information on floods in the upstream is not transmitted to downstream settlements; hence insufficient warning is supplied for any pre- activities to operate flood protection devices (Chen et al., 2017).

Many scholars have done FPM and natural risk assessment using remote sensing (RS) and geographic information systems (GIS) globally, attempting to make a substantial contribution to flood risk assessment (Abdelkader et al., 2013; Chatterjee et al., 2003). Using either statistical or deterministic algorithms, FPM can reliably detect and characterize future flood risk. Furthermore, flood probability of a region can be assessed qualitatively or quantitatively (Sahana and Sajjad, 2019; Sinha et al., 2008). The three distinct categories of geospatial techniques utilized in flood analysis are as follows: hydrologic or hydrodynamic models (Herder, 2013), statistical and multiple criteria decision making (MCDM) methods (Ayalew and Yamagishi, 2005), and machine learning algorithms (MLAs) (Arabameri et al., 2020) However, the aforementioned techniques are not devoid of flaws in deriving a flood probability map. For example, results of the MCDM model are subject to disparities as a result of biased expert judgment (Ghorbanzadeh et al., 2019; Paquette and Lowry, 2012). Moreover, results of the statistical models depend heavily on the sample size (Liao and Carin, 2009).On the other hand, hydrodynamic models convert discharge flows into flood depths or flood velocity (Al-Mulali et al., 2015). Although this model yields relatively accurate results for small basin, it is challenging to apply this model to entire area (Guo et al., 2012; van Emmerik et al., 2015). MLAs with competence in nonlinear mapping have been extensively developed in the past few decades as alternatives to address these shortcomings (Huang et al., 2019)

A number of studies have been conducted on floods in Nepal. These discussions have emphasized flood risks (Karki et al., 2011), mitigation measures (Devkota et al., 2014), flood forecasting (Gautam and Phaiju, 2013), and river hydrology or morphology (Marston et al., 1996). Although a lot of studies have been dedicated to improving reliability and accuracy of flood probability mapping to-date, there has been no algorithm that can achieve the best accuracy for all areas or regions. In some cases, an algorithm that outperforms other algorithms in one area may not perform well in another. This may be caused by complex physical processes that vary from region to region and directly impact flood influencing factors. Because of this, scholars need to continually assess the reliability of newly developed algorithms in flood probability modeling. This study investigates the potential of multiple kernel

functions in machine learning algorithms to improve precision in flood probability mapping at a basin scale. Because there is a lack of precise flood maps, this study is expected to contribute significantly to flood prediction and subsequent recommendations for avoiding human injuries, deaths, and property damage.

This study addresses the gap by examining relevant flood probability modeling literature, followed by a description of the materials and methods in **Section 2**. **Section 3** and **Section 4**, respectively, describes results and discussion. Finally, **Section 5** concludes major findings of this work with future research directions.

2. Materials and methods

2.1 Description of the study area

Nepal is a land lock country, situated in South Asia with a total 147,181 km^2 area between 26º22′ to 30º27′ N latitudes and 80º04′ to 88º12′ E longitudes (Shrestha, 2007). The whole country is divided into three major physiographic regions, i.e., mountains, hills, and low land (Tarai). These regions have their importance in water, crop, and industrial production.

Dhanusha and Mahottari, with a total area of 2182 km2 between latitude 26°50′31.56″ North, longitude 86°02′09.60″ East have been considered for the case study. These districts are administratively placed in Province two of the Nepal federal government.

The districts are in the Koshi basin belt. Following the classification by Hagen, (1969), these districts have been characterized by numerous physiographic regions. Upper areas are affiliated with Shiwalik (Chure) hills and Dun valleys. In contrast, southern parts are alluvial deposited (Terai) flat land, which is the northern part of the Indo-Gangetic Plain. The two distinct physiographic regions, such as Shiwalik hill and Terai land, are tectonically separated by Main Frontal Thrust (MFT) and Main Boundary Thrust (MBT). The climate is characterized by tropical climate that primarily depends on the Asian monsoon rainfall, and approximately 80% of the annual rainfall occurs from June to September (Sharma et al., 2019). The climate becomes hot from April to September, while it is cold from December to March. Climate has a significant influence on the vegetation processes; for example, tropical forest species are distributed in the Shiwalik region whereas the Terai land is demonstrated for the agricultural practices. Roads are primary means of transportation in the study area as the majority of the road are earthen which quickly get inundated during the monsoon period.

The study area falls in the arid and semi-arid hydrological zones, having average rainfall of 1,480 mm to 2,200 mm and an average temperature range between 11°C and 40°C (Joshi and Dongol, 2018).

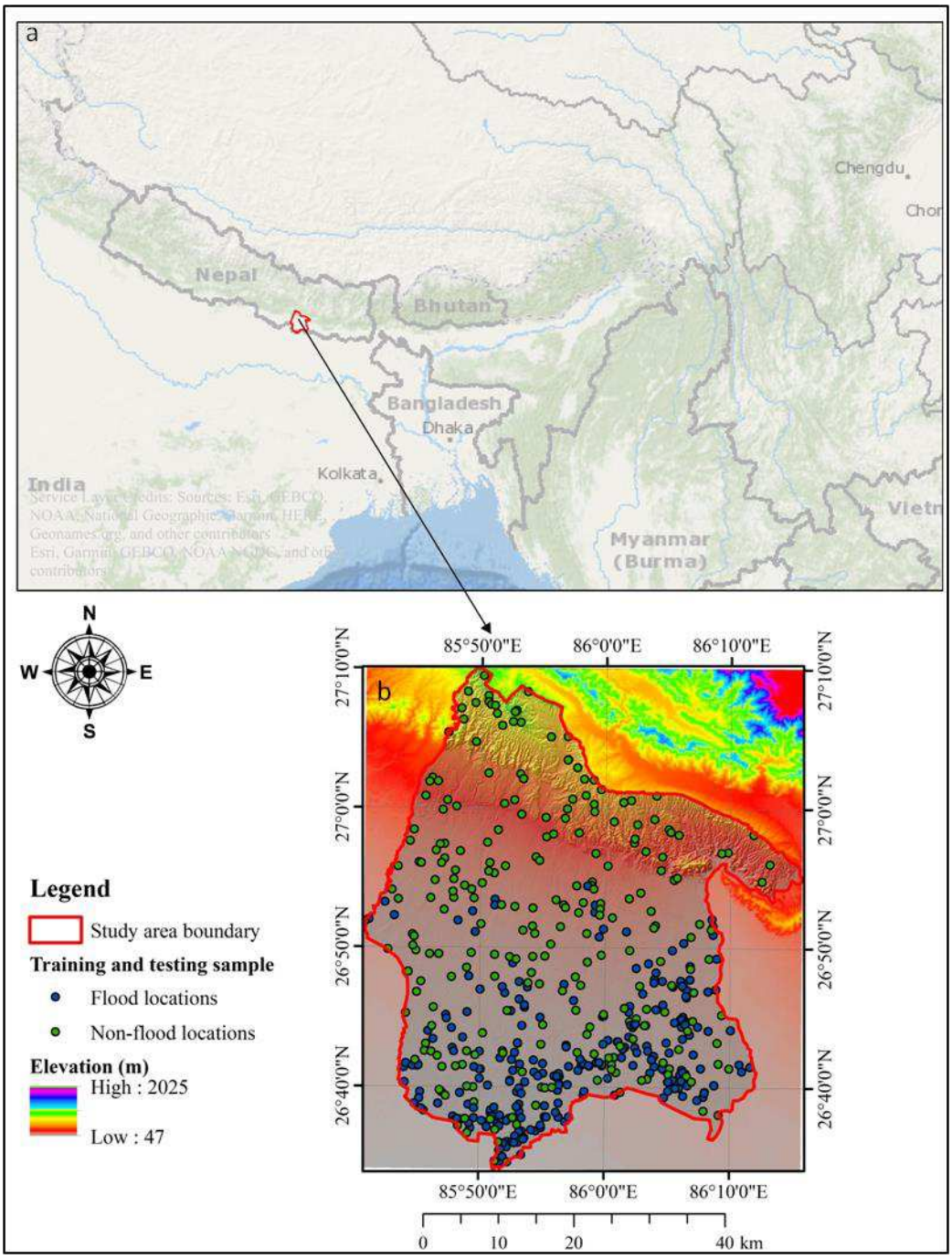


Fig. 1 Study area with flood locations.

2.2 Modelling flood probability

2.2.1 Developing flood inventory map

The flood inventory map, which depicts historical flood locations, is prepared using the 250 flood locations as flood points found in the existing flood reports, historical datasets between 2000 and 2020. , Landsat 5-8 (https://earthexplorer.usgs.gov/), and Sentinel 1 (https://search.asf.alaska.edu/) satellite images (Fig. 1). In addition, there were also 250 locations defined as non-flood points. Afterwards, flood and non-flood points were split into training and testing subsets (70:30) using a random selection process in a GIS environment (Rahmati and Pourghasemi, 2017; Samanta et al., 2018; Tehrany and Kumar, 2018). Precisely, 350 (70%) flood locations were used to train the algorithms, and 150 locations were used for test the algorithms.

2.2.2 Factors contributing to floods

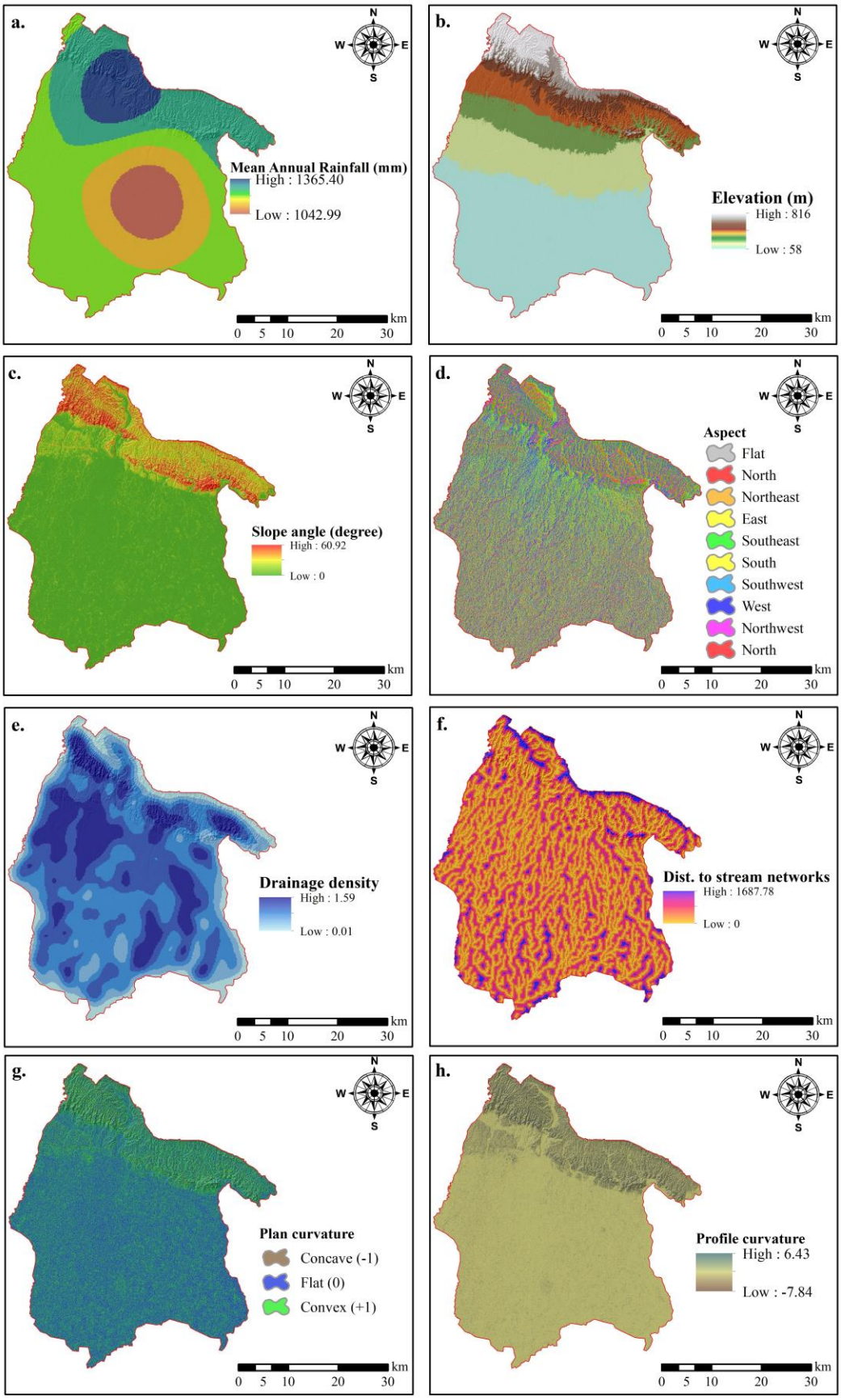
To build a reliable flood probability model, it is important to identify the most influencial contributing factors (Bui et al., 2019). Concerning local hydro-environmental characteristics, the effect of contributing factors to flooding varies from one space to another (Cao et al., 2016). According to the literature and available data, there were thirteen contributing factors: rainfall, elevation, slope angle, aspect, drainage density (DD), distance to stream networks, plan curvature, profile curvature, land cover, lithology, soil texture, steam power index (SPI), and topographic wetness index (TWI) considered for this study (Bui et al., 2018; Choubin et al., 2019; Khosravi et al., 2018; Wang and Xu, 2017).

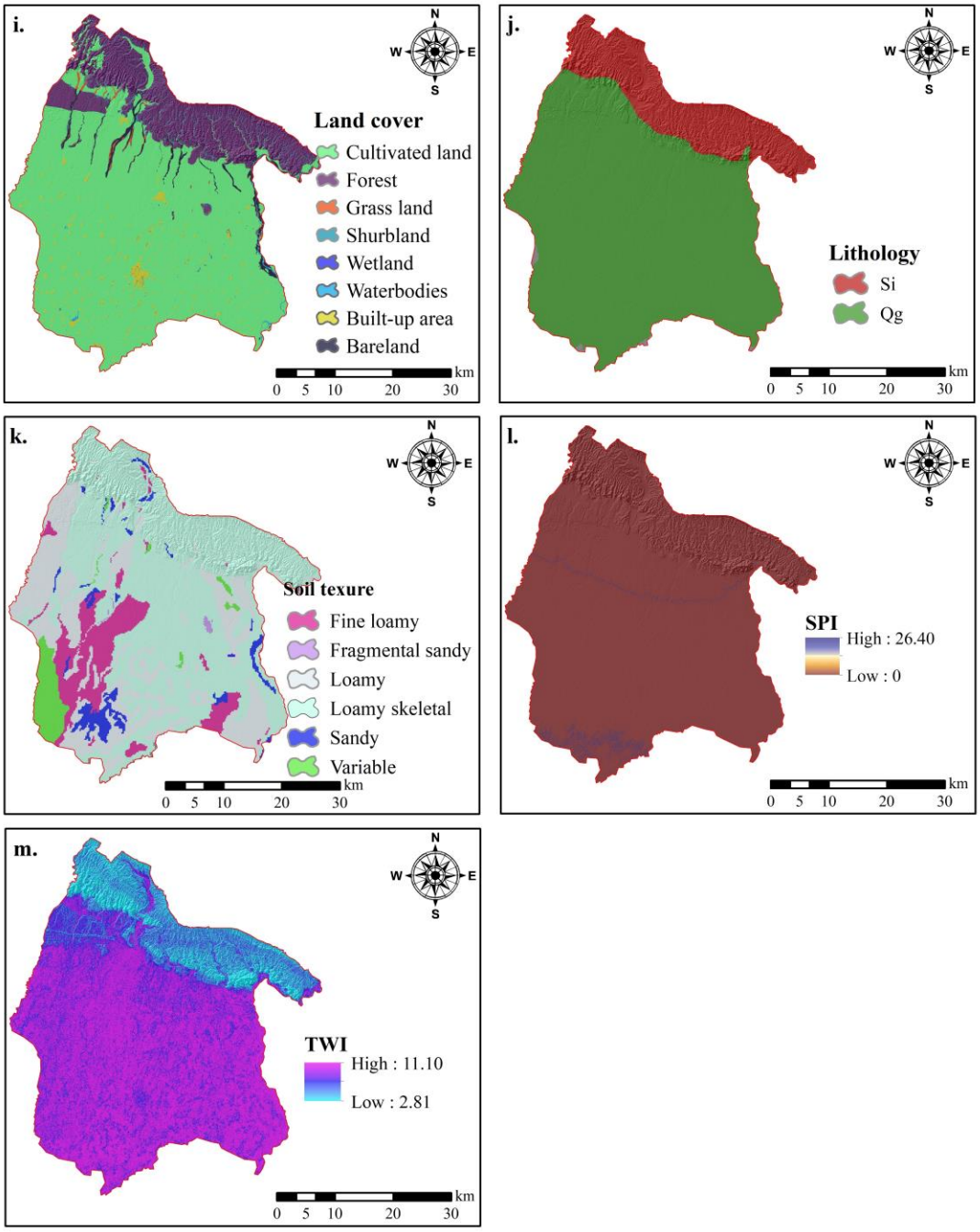
A digital elevation model (DEM) with a resolution of 30 m was acquired from advanced land observation satellite (ALOS; https://www.earthdata.nasa.gov/). Later, DEM data was used to extract topographic factor maps, i.e., elevation, slope angle, aspect, DD, distance to stream networks, plan and profile curvatures, SPI, and TWI. A rainfall map was constructed from rainfall data for the periods between 1990 and 2020, obtained from Department of Hydrology and Meteorology of Nepal (https://www.dhm.gov.np/) using the kriging interpolation approach with four data points (Shekhar and Pandey, 2015; Szwagrzyk et al., 2018). According to the literature, kriging is the most appropriate approach in data-sparse country (Arabameri et al., 2020; Chowdhuri et al., 2020). A soil map was obtained from the land resource mapping project (LRMO 1986; http://opac.narc.gov.np/), while the lithology map was found from the Department of Mines and Geology of Nepal (https://www.dmgnepal.gov.np/). A land cover map was acquired from the GLOBALLAND30 (https://www.globallandcover.com). Note that thematic layers of

contributing factors were resampled to a 30 m resolution to make the same grid size of all layers. Further, flood contributing factors were reclassified. Details on flood contributing factors can be found in Fig. 2 and Table 1.

Table 1 Thematic layers of contributing factors used in this study.

Classification	Sub-classification/Information	References	Resolution/Scale
Flood inventory	Historical flood marks, Landsat 5-8, and Sentinel-1	https://earthexplorer.usgs.gov https://search.asf.alaska.edu	30 m × 30 m
DEM	Elevation, slope angle, aspect, drainage density, distance to stream networks, plan and profile curvatures, SPI, and TWI	https://www.earthdata.nasa.gov	30 m × 30 m
Mean annual rainfall	Historical data (1990-2020)	https://www.dhm.gov.np/	30 m × 30 m
Land cover	GLOBALLAND30	https://www.globallandcover.com	30 m × 30 m
Soil texture	Land resource mapping project (LRMO 1986	http://opac.narc.gov.np	1: 50,000
Lithology	Department of Mines and Geology of Nepal	https://www.dmgnepal.gov.np	1: 50,000


a.
N
W
E
S
Mean Annual Rainfall (mm)
High : 1365.40
Low : 1042.99
km
0 5 10 20 30
b.
Elevation (m)
High : 816
Low : 58
c.
Slope angle (degree)
High : 60.92
Low : 0
d.
Aspect
Flat
North
Northeast
East
Southeast
South
Southwest
West
Northwest
North
e.
Drainage density
High : 1.59
Low : 0.01
f.
Dist. to stream networks
High : 1687.78
Low : 0
g.
Plan curvature
Concave (-1)
Flat (0)
Convex (+1)
h.
Profile curvature
High : 6.43
Low : -7.84



138 **Fig. 2** Flood contributing factors: (a) mean annual rainfall, (b) elevation, (c) slope angle, (d)
139 aspect, (e) drainage density, (f) distance to stream networks, (g) plan curvature, (h) profile
140 curvature, (i) land cover, (j) lithology, (k) soil texture, (l) SPI, and (m) TWI.

141 2.2.3 Multicollinearity analysis

Factors that are linearly related with other factors are referred to as multicollinearity. It results in somewhat redundant variables. In this study, multicollinearity was measured with the variance inflation factor (VIF) that evaluates how much an expected regression coefficient variance tends to increase, if certain predictors are correlated (Arabameri et al., 2020; Dormann et al., 2013; Wang et al., 2020). In addition, tolerance (TOL) can also help in detecting multicollinearity. Multicollinearity is problematic if the TOL value is less than 0.1 and simultaneously the VIF 5 and higher (Kutner et al., 2005; Roy et al., 2020). In this case, the redundant factor/s should be eliminated from the algorithm.

2.2.4 Feature selection

The Pearson coefficient was computed to consider valuable features in the algorithm. This study evaluated the coefficient threshold as < 0.70 to use the feature during algorithm building. This means if the variables are associated (≥ 0.70), then one of them must be retained and the rest be dropped (Dormann et al., 2013).

2.2.5 Spatial modeling

The Gaussian process regression (GPR) usually employs translation-invariant covariances, resulting in a highly versatile and nonlinear prediction mechanism (Blix et al., 2017). It can detect information in any kernel-based regression process, compute quickly, and express this in a closed-form. It specifies regressor learning within a Bayesian system, suggesting the model variables perform a Gaussian distribution interpreting prior information of the final output (Colkesen et al., 2016). GPR was recently implemented and utilized in numerous machine learning applications (Colkesen et al., 2016; Zhao et al., 2011). Hence, this study considered GPR as one of the algorithms to model flood probability. The reader is directed to Kuss (2006), Sihag et al. (2018), and Paananen et al. (2019) for additional details on GPR and various covariance functions.

The support vector machine (SVM) is a supervised technique, based on mathematical learning theory and systemic risk minimization (Vapnik, 1995). Boser et al. (1992) introduced the idea of SVM, which is one of the most powerful tools for flood probability mapping. The SVM algorithm intends to find an optimum separating hyperplane that can differentiate between the two categories: floods and non-floods with the training dataset (Huang and Zhao, 2018). Precisely, SVM is a learning tool used for classification and regression to eliminate errors in classification or fitness functions (Sihag et al., 2018). Considering its versatility in dealing with multi-dimensional datasets and superior generalization performance, SVM has been primarily used to model flood probability in this study (Yang

and Cervone, 2019). Detailed descriptions of this algorithm can be found in Ren et al. (2015), Sachdeva et al. (2017), and Band et al. (2020).

In addition, this study was designed based on multiple kernel functions in both SVM and GPR, e.g., Pearson VII function kernel (PUK), polynomial, normalized poly kernel, and radial basis kernel function (RBF). For details regarding kernel functions, the reader refers to Xing et al. (2016), Sihag et al. (2017), and Sihag et al. (2018). Table 2 depicts hyper-parameters of GRP and SVM algorithms used in this study.

Table 2 Hyper-parameters of machine learning algorithms (GPR and SVM) used in this study

Model name	Description of parameters
Gaussian Process Regression (GPR)	Kernel=Normalized Poly; Batch size-100, Noise = 1, Seed = 1; Filter type =Normalize training data; Cache size = 250000 Kernel= Poly; Batch size-100, Noise = 1, Seed = 1; Filter type =Normalize training data; Cache size = 250000 Kernel= PUK; Batch size-100, Nose = 1, Seed = 1; Filter type =Normalize training data; Cache size = 250000; Omega = 0.01; Sigma =0.01 Kernel= RBF; Batch size-100, Noise = 1, Seed = 1; Filter type =Normalize training data; Cache size = 250000; Gamma = 0.01
Support Vector Machine (SVM)	Kernel=Normalized Poly; Batch size=100, C= 1,Regression Optimizer = SMO Improved; Filter type =Normalize training data; Cache size = 250000 Kernel= Poly; Batch size=100, C= 1, Regression Optimizer = SMO Improved; Filter type =Normalize training data; Cache size = 250000 Kernel= PUK; Batch size=100, C= 1, Regression Optimizer = SMO Improved; Filter type =Normalize training data; Cache size = 250000; Omega = 0.01; Sigma = 0.01 Kernel= RBF; Batch size=100, C= 1, Regression Optimizer = SMO Improved; Filter type =Normalize training data; Cache size = 250000; Gamma = 0.01

2.2.6 Model performance assessment

Multiple statistical indices, i.e., coefficient of determination (r^2), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root-relative square error (RRSE) were used to check the goodness-of-the-fit (training datasets) and performance (testing datasets) (Mohammadzadeh et al., 2014) of our proposed models. These indices can be computed using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(M_{predicted} - M_{observed}\right)^2}{\sum_{i=1}^{N}(M_{observed} - \bar{M}_{observed})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(M_{predicted} - M_{observed}\right)\right]^2} \quad (2)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|M_{predicted} - M_{observed}\right| \quad (3)$$

$$RAE = \frac{\left[\sum_{i=1}^{N}\left(M_{predicted} - M_{observed}\right)^2\right]^{1/2}}{\left[\sum_{i=1}^{N} M_{observed}^2\right]^{1/2}} \quad (4)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{N}\left(M_{predicted} - M_{observed}\right)^2}{\sum_{i=1}^{N}(M_{observed} - \bar{M}_{observed})^2}} \quad (5)$$

where, N denotes the number of observations; $M_{predicted}$ is the predicted value; $M_{observed}$ is actual value; and $\bar{M}_{observed}$ is the mean of actual values.

2.2.7 Model validation

In this study, flood data of 2020, sourced from the International Centre for Integrated Mountain Development (ICIMOD), were used to validate the model (https://doi.org/10.26066/RDS.36038). To this end, we have considered 40 locations (20 flood and 20 non-flood points) from the flood inventory. A confusion matrix, including sensitivity, specificity, and accuracy, was computed to validate the flood probability map (Khosravi et al., 2018). Equations used to validate best-fitted algorithm are presented as:

$$Sensitivity = \frac{True\ positive\ (TP)}{True\ positive\ (TP) + False\ negative\ (FN)} \quad (6)$$

$$Specificity = \frac{True\ negative\ (TN)}{True\ negative\ (TN) + False\ positive\ (FP)} \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

3. Results

3.1 Multicollinearity and feature selection

Linear correlation between features induces multicollinearity problem, leading to algorithm inconsistency and feature data redundancy (Katrutsa and Strijov, 2017). Except for slope angle and lithology, multicollinearity test results have no multicollinearity issue observed (TOL > 0.1 and VIF <5; Table 3). Likewise, in feature selection,

194 results showed the same findings. Slope angle and lithology have crossed the threshold value (≥ 0.70; Fig. 3a).
195 Taking this into account, we have eliminated two redundant factors from the algorithm and rechecked the featured
196 threshold value, and found that no factors crossed the assigned threshold (< 0.70; Fig. 3b). Therefore, eleven factors,
197 including rainfall, elevation, aspect, drainage density, distance to stream networks, plan and profile curvatures, LC,
198 soil texture, SPI, and TWI, were further considered for building the algorithms in flood probability modeling.

199 **Table 3** Multicollinearity statistics (LC: land cover, SPI: Stream Power Index, and TWI:
200 Topographic Wetness Index)

Factor	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Rainfall	-0.013	0.028	-0.024	-0.474	0.636	0.557	1.795
Elevation	-0.145	0.026	-0.476	-5.652	0.000	0.203	4.917
Slope angle	0.066	0.039	0.148	1.689	0.092	0.189	5.282★
Aspect	-0.016	0.008	-0.078	-2.017	0.044	0.955	1.047
Drainage density	-0.052	0.018	-0.120	-2.935	0.003	0.860	1.163
Distance to stream networks	-0.024	0.021	-0.047	-1.163	0.246	0.884	1.131
Plan Curvature	-0.021	0.041	-0.021	-0.505	0.614	0.818	1.222
Profile Curvature	-0.032	0.037	-0.037	-0.863	0.389	0.795	1.257
LC	-0.001	0.001	-0.023	-0.572	0.567	0.886	1.129
Lithology	-0.382	0.153	-0.229	-2.497	0.013	0.172	5.819★
Soil Texture	0.026	0.006	0.183	4.305	0.000	0.799	1.252
SPI	0.046	0.076	0.024	0.608	0.543	0.957	1.045
TWI	0.126	0.035	0.274	3.553	0.000	0.243	4.115

★**Note:** Shaded red marks indicate the sign of collinearity and should be eliminated from the model

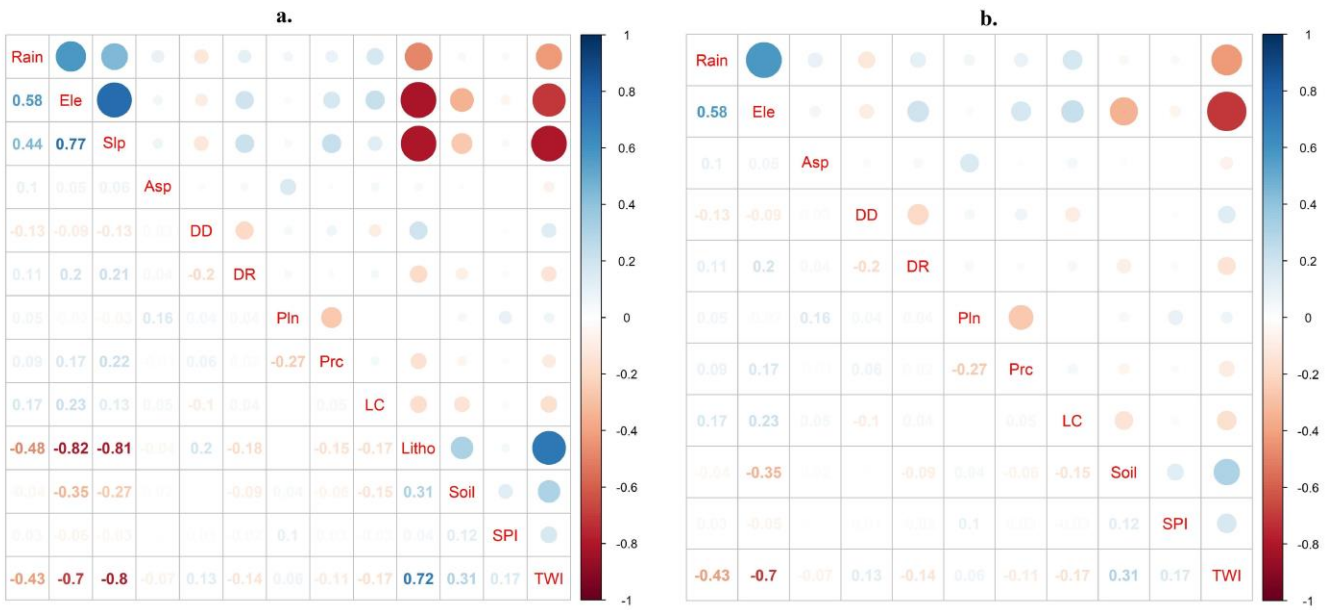


Fig. 3 Feature selection: (a) total features and (b) eliminated features due to higher correlation using Pearson's correlation analysis. (Rain, rainfall; Ele, elevation; Slp, slope angle; Asp, aspect; DD, drainage density; Pln, plan curvature; Prc, profile curvature; LC, land cover; Litho, lithology; Soil, soil texture; SPI, stream power index; and TWI, topographic wetness index)

3.2 Goodness-of-the-fit and model performance assessment

The overall performance, including goodness-of-the-fit (with training datasets) and predictive capability (with testing datasets) of the algorithms, is calculated via the r^2, MAE, RMSE, RAE, and RRSE (Table 4) and compared the algorithms to verify which of them are the most predicting capability. In the training and testing phase, The SVM-PUK kernel (0.990 and 0.995) has the highest r^2, followed by the GPR-PUK kernel (0.973 and 0.989), the GPR-Normalized Poly kernel (0.585 and 0.603), the SVM-Normalized Poly kernel (0.562 and 0.606), the GPR-Poly kernel (0.534 and 0.563), the GPR-RBF kernel (0.528 and 0.549), the SVM-RBF kernel (0.520 and 0.546), and the SVM-Poly kernel (0.515 and 0.546); these indicate except for SVM-PUK kernel, and GPR-PUK kernel could not be capable of explaining the entire spectrum of variability in the output performance. As a result, it is critical to employ a diverse set of output parameters in order to classify the best algorithms more reliably. On the other hand, the MAE, RMSE, RAE, and RRSE demonstrate the highest efficiency of the SVM-PUK kernel algorithm for both training and testing datasets in predicting floods (cf. Section 3.3).

Table 4 Statistical indices for model goodness-of-the-fit and performance assessment

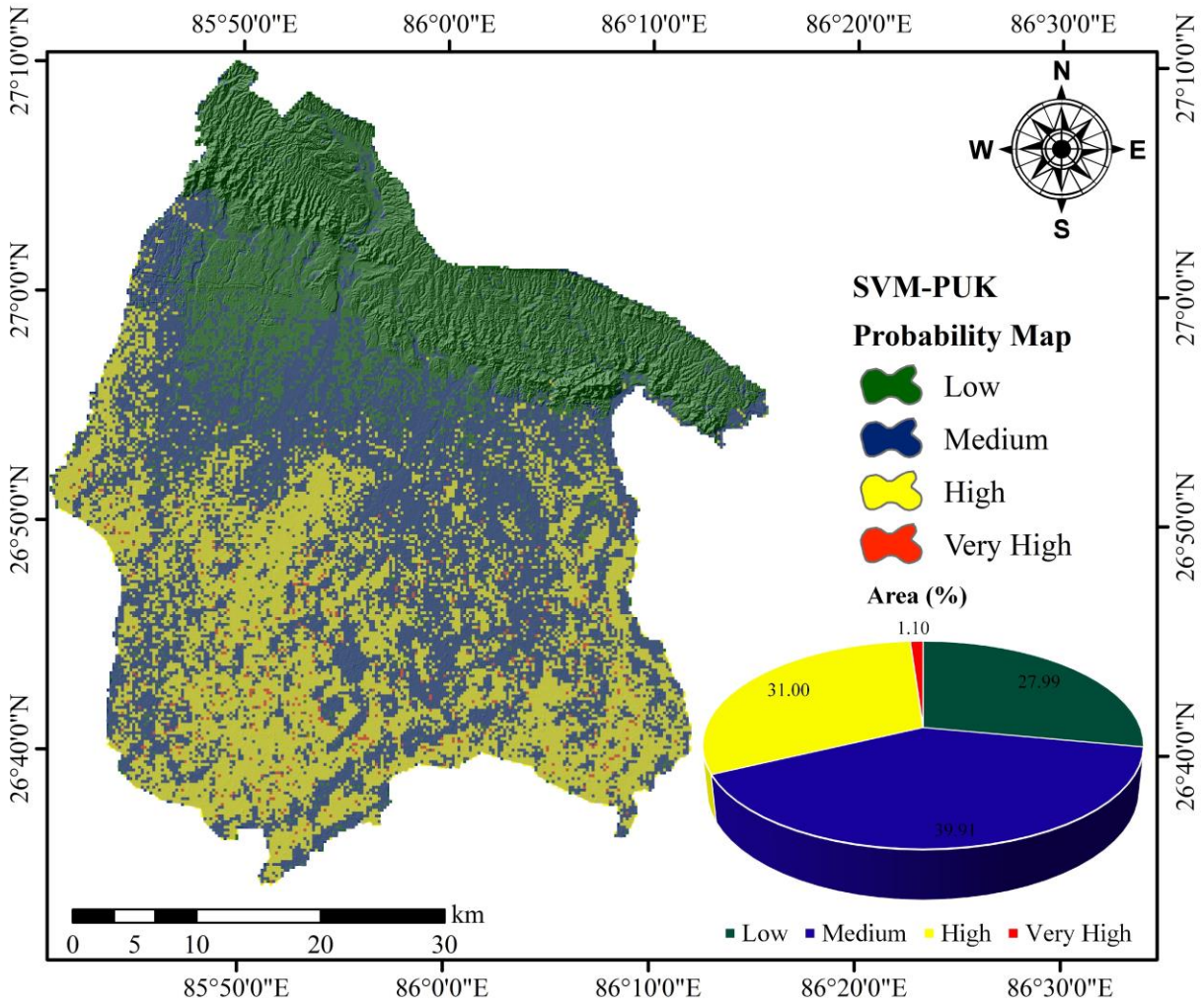
Training dataset								
Index	**GPR-PUK kernel**	**GPR-Poly kernel**	**GPR-Normalized Poly kernel**	**GPR-RBF kernel**	**SVM-PUK kernel**	**SVM-Poly kernel**	**SVM-Normalized Poly kernel**	**SVM-RBF kernel**
R^2	0.973	0.534	0.585	0.528	0.990	0.515	0.562	0.520
MAE	0.299	0.381	0.358	0.416	0.012	0.352	0.303	0.356
RMSE	0.303	0.423	0.406	0.437	0.071	0.448	0.435	0.450
RAE (%)	59.80	76.25	71.52	83.11	3.32	70.46	60.58	71.26
RRSE (%)	60.55	84.56	81.28	87.45	14.20	89.56	87.03	89.91
Testing dataset								
Index	**GPR-PUK kernel**	**GPR-Poly kernel**	**GPR-Normalized Poly kernel**	**GPR-RBF kernel**	**SVM-PUK kernel**	**SVM-Poly kernel**	**SVM-Normalized Poly kernel**	**SVM-RBF kernel**
R^2	0.989	0.563	0.603	0.549	0.995	0.546	0.606	0.546
MAE	0.312	0.373	0.351	0.443	0.009	0.342	0.306	0.367
RMSE	0.313	0.413	0.401	0.451	0.051	0.466	0.422	0.455
RAE (%)	62.30	74.52	70.21	88.55	1.81	68.46	61.14	73.47
RRSE (%)	62.55	82.67	80.12	90.25	10.19	93.23	84.31	91.00

Note: Green shaded marks indicate the best-fitted model for training and testing datasets

3.3 Flood probability mapping

We affirmed that the SVM-PUK kernel algorithm is the most suitable and reliable algorithm for flood modeling. The flood probability map is displayed in Fig. 4. The following four hazardous classes were assigned to the flood probability map: low, medium, high, and very high using the natural break classification method. The best algorithm (SVM-PUK kernel) results show that 27.99%, 39.91%, 31.00%, and 1.10% of the total land areas comprise low, medium, high, and very high flood probabilities, respectively.

According to our results, floods mainly occur adjacent to the riverside. In addition, low-lying areas are most prone to flood hazards. Specifically, probability of floods in the downstream regions of the southwestern part of the study area is more prone to floods than in the southeastern part.



227 **Fig. 4** Flood probability map using the best-fitted algorithm (SVM-PUK kernel)

228 3.4 Validation of flood probability map

229 The resultant flood map was validated using the 2020 flood inventory data. According to results, an accuracy value
230 of 0.88 indicated that the algorithm almost perfectly classified 88.0% of the flood points. Therefore, it can be
231 concluded that the algorithm output (SVM-PUK kernel) was compatible with the observed inventory data. The
232 results are presented in Table 5.

233 **Table 5** Validating flood probability map based on 2020 flood inventory data

	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
Values	18	17	2	3	0.86	0.89	0.88

234 **4. Discussion**

Given extreme weather events such as flooding could enhance in the days ahead damage to infrastructures may swell, which is likely to increase economic losses Hence it is critical to developing an approach for calculating associated significant socioeconomic losses. The ability to generate cost-effective flood modeling would benefit relevant authorities of performing disaster management activities. This study offered and evaluated a basin-scale flood modeling with multiple innovative machine learning algorithms (MLAs). Comparing flood probability models using multiple MLAs is problematic because it necessitates understanding the complex relationships between floods and their influential factors (Arabameri et al., 2020; Youssef et al., 2015). However, the performance of the MLAs can be portrayed as optimum concerning time, cost and efficiency (Lee et al., 2017). In addition, these approaches could be very valuable in providing speedy flood mapping in emergency scenarios comparing traditional flood modeling, i.e., hydrological and hydrodynamical modeling (Chau and Lee, 1991). Nonetheless, significant room exists for improvement in order to achieve superior results (Fenicia et al., 2014).

The 175 floods and 175 non-flood locations were randomly split to test the algorithms for thirteen contributing factors. We then evaluated multicollinearity (TOL and VIF) among floods and their contributing factors. The multicollinearity results exhibited slope angle and lithology have the same influence on flooding as other factors. Therefore, these two factors were excluded from the algorithm before the final training. The findings are inlined with those studies refer to(Arabameri et al., 2020; Dormann et al., 2013; Shahin and Hassan, 2000). In addition, we evaluated important features in machine learning modeling using feature selection techniques such as the Pearson correlation coefficient. According to results of this study (threshold must be < 0.70), slope angle and lithology had higher coefficient value (> 0.70), therefore, these features were regarded as redundant and excluded from algorithm training. The studies by (Chen et al., 2020; Hong et al., 2018; Miles, 2014; Xu and Li, 2020) reported a similar observation and also eliminated the redundant features.

It is worth noting that highest accuracy of the algorithms applied in this study, obtained by the SVM-PUK kernel (r^2: 0.995, Accuracy: 0.88), exceeded the precision of the algorithm proposed by (Sahana and Patel, 2019), which constructed flood probability map within Koshi River basin (lower part: India). In addition, the GPR-PUK kernel algorithm (r^2: 0.973) proposed in this study obtained higher precision compared with the models offered by (Sahana and Patel, 2019). As the proposed algorithms are new to flood probability modeling and no other studies were found in the literature regarding flood mapping with these algorithms; hence, we could not compare the proposed

algorithms with other similar studies. However, accuracy of the proposed algorithm in constructing a flood probability map will give significant insight into the scientific community.

The Koshi River is a mighty river system with a history of reversing course and wreaking havoc in Nepal (Devkota et al., 2012). Thousands of families persist in fear that the Koshi River would perhaps burst its banks at any moment, creating widespread hardship (Khanal et al., 2007). Moreover, this river has a proven history of changing river paths, making it difficult and dynamic to address the risk of floods (Tiwari and Joshi, 2012). During the monsoon season, the large volume of upstream flow affects downstream catchments and the floodwaters in the southeast and southwest of the area. Apart from these main causes, river density is critical in defining more and less flooded areas and high dense regions created by porous soil and low elevated regions (Mukerji et al., 2009). The dense and abundant rainfall is observed in the downstream portion, which has a greater risk of flooding than in other study regions. It has been found in previous research that dense and abundant rainfall has a greater flood probability (Donat et al., 2013; Sharma et al., 2000; You et al., 2008). It is worth remembering that the Koshi River flood is a clear signal of Nepal's flood susceptibility and its need to tackle this susceptibility by reducing and preparing catastrophe risks rather than concentrating on humanitarian support.

In this study, we acquired flood contributing factors from various sources and filtered them for relevance. ALOS DEM products are widely available and easy to use and have been commonly used in many studies to extract flood contributing factors https://www.earthdata.nasa.gov. In addition, in hydrological analysis, some advanced DEM products, i.e., SRTM (30 m) and ALOS PALSAR (12.5 m) DEM outperformed ALOS DEM. Therefore, the effect of various advanced DEM products on flood probability modeling is needed. Furthermore, it is essential to note that the main shortcomings of this study are a temporal discrepancy between flooding and its contributing factors. Lithology and soil texture can generally be considered constant and do not change over the years.

Nevertheless, seasonal changes will affect land cover, DEM, and DEM derive factors. Moreover, particular flood contributing factors can change significantly within each year. Hence, developing a dynamic analysis of flood probability with different temporal aspects will become increasingly important in the future work.

Lastly, it should be noted that using various machine learning algorithms and geospatial techniques is extremely effective in conducting FPM based on time, expenses, and precision without expert judgment in modeling. This research also provided insights into the execution of FPM in basin-scale since the findings achieved are relevant to

regional and local governments of flood-prone areas, which shows that FPM can be effectively executed in the other region with similar environmental characteristics.

5. Conclusions

This study evaluated innovative multiple kernel functions in the machine learning algorithms, i.e., GPR and SVM with GPR-PUK, GPR-Poly, GPR-Normalized Poly, GPR-RBF, SVM-PUK, SVM-Poly, SVM-Normalized Poly, and SVM-RBF kernels in flood probability mapping. The proposed algorithms have not been examined previously and are presented here to exploit the advantages of multiple kernel functions for basin-scale flood modeling. Results suggested that the SVM-PUK kernel algorithm performed the best, with SVM-PUK R^2: 0.990 and 0.995, SVM-PUK MAE: 0.012 and 0.009, SVM-PUK $RMSE$: 0.071 and 0.051, SVM-PUK RAE: 3.32 and 1.81%, and SVM-PUK $RRSE$: 14.20 and 10.19% for training and testing datasets than other algorithms, which is a significant success in the scope of the study. Besides, the algorithm's accuracy (88%) confirmed that it could almost perfectly classify the flood points in the study area. Therefore, the findings of this study can be helpful to relevant authorities' direct flood control initiatives in basin-scale to minimize vulnerability to flooding, enhance flood early warning systems, and evacuate flood victims.

Dynamic influence of climatic and environmental variables, e.g., rainfall and land cover, were not considered in this study. Future research is recommended to consider climatic and land cover dynamic characteristics to predict flood probability. In addition, future studies will consider multiple climate models combined with MLAs to predict future flood scenarios in the study area.

CRediT author statement

Baig, M.A., Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Xiong, D.H.,** Investigation, Writing-review and editing, Supervision, Funding acquisition. **Rahman, M.,** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft. **Islam, M.M.,** Investigation, Writing-review and editing. **Elbeltagi, A.,** Software, Writing-review and editing. **Yigez, B.,** Writing-review and editing. **Rai, D.K.,** Writing-review and editing. **Tayab, M.,** Writing-review and editing. **Dewan, A.,** Investigation, Writing-review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the CAS, Chinese Academy of Sciences Overseas Institutions Platform Project (Grant No. 131C11KYSB20200033). The authors acknowledge and appreciate the provision of rainfall data by the Department of Hydrology and Meteorology of Nepal, without which this study would not have been possible.

Data Availability Statement

Data are available upon request on the corresponding author.

References

Abdelkader, M., Shaqura, M., Claudel, C. G., and Gueaieb, W., A UAV based system for real time flash flood monitoring in desert environments using Lagrangian microsensors, *in* Proceedings 2013 International Conference on Unmanned Aircraft Systems (ICUAS)2013, IEEE, p. 25-34.

Al-Mulali, U., Ozturk, I., and Lean, H. H., 2015, The influence of economic growth, urbanization, trade openness, financial development, and renewable energy on pollution in Europe: Natural Hazards, v. 79, no. 1, p. 621-644.

Arabameri, A., Chen, W., Loche, M., Zhao, X., Li, Y., Lombardo, L., Cerda, A., Pradhan, B., and Bui, D. T., 2020, Comparison of machine learning models for gully erosion susceptibility mapping: Geoscience Frontiers, v. 11, no. 5, p. 1609-1620.

Ayalew, L., and Yamagishi, H., 2005, The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan: Geomorphology, v. 65, no. 1-2, p. 15-31.

Band, S. S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Melesse, A. M., and Mosavi, A., 2020, Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms: Remote Sensing, v. 12, no. 21, p. 3568.

Blix, K., Camps-Valls, G., and Jenssen, R., 2017, Gaussian process sensitivity analysis for oceanic chlorophyll estimation: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 10, no. 4, p. 1265-1277.

Boser, B. E., Guyon, I. M., and Vapnik, V. N., A training algorithm for optimal margin classifiers, *in* Proceedings Proceedings of the fifth annual workshop on Computational learning theory1992, p. 144-152.

Bui, D. T., Ngo, P.-T. T., Pham, T. D., Jaafari, A., Minh, N. Q., Hoa, P. V., and Samui, P., 2019, A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping: Catena, v. 179, p. 184-196.

Bui, D. T., Panahi, M., Shahabi, H., Singh, V. P., Shirzadi, A., Chapi, K., Khosravi, K., Chen, W., Panahi, S., and Li, S., 2018, Novel hybrid evolutionary algorithms for spatial prediction of floods: Scientific reports, v. 8, no. 1, p. 1-14.

Cao, C., Xu, P., Wang, Y., Chen, J., Zheng, L., and Niu, C., 2016, Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas: Sustainability, v. 8, no. 9, p. 948.

Chatterjee, C., Kumar, R., and Mani, P., 2003, Delineation of surface waterlogged areas in parts of Bihar using IRS-1C LISS-III data: Journal of the Indian Society of Remote Sensing, v. 31, no. 1, p. 57-65.

Chau, K. W., and Lee, J., 1991, Mathematical modelling of Shing Mun river network: Advances in Water Resources, v. 14, no. 3, p. 106-112.

Chen, N., Hu, G., Deng, W., Khanal, N. R., Zhu, Y., and Han, D., 2017, Water Hazards in the Trans-boundary Kosi River Basin, Land Cover Change and Its Eco-environmental Responses in Nepal, Springer, p. 383-408.

Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., and Pradhan, B., 2020, Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods: Science of The Total Environment, v. 701, p. 134979.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., and Mosavi, A., 2019, An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines: Science of the Total Environment, v. 651, p. 2087-2096.

Chowdhuri, I., Pal, S. C., and Chakrabortty, R., 2020, Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India: Advances in Space Research, v. 65, no. 5, p. 1466-1489.

Colkesen, I., Sahin, E. K., and Kavzoglu, T., 2016, Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression: Journal of African Earth Sciences, v. 118, p. 53-64.

Dano, U. L., Balogun, A.-L., Matori, A.-N., Wan Yusouf, K., Abubakar, I. R., Said Mohamed, M. A., Aina, Y. A., and Pradhan, B., 2019, Flood susceptibility mapping using GIS-based analytic network process: A case study of Perlis, Malaysia: Water, v. 11, no. 3, p. 615.

Devkota, L., Crosato, A., and Giri, S., 2012, Effect of the barrage and embankments on flooding and channel avulsion case study Koshi River, Nepal: Rural Infrastructure 3 (3), 124-132.(2012).

Devkota, R. P., Cockfield, G., and Maraseni, T. N., 2014, Perceived community-based flood adaptation strategies under climate change in Nepal: International Journal of Global Warming, v. 6, no. 1, p. 113-124.

Dixit, A., 2009, Kosi embankment breach in Nepal: Need for a paradigm shift in responding to floods: Economic and political weekly, p. 70-78.

Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., and Caesar, J., 2013, Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset: Journal of Geophysical Research: Atmospheres, v. 118, no. 5, p. 2098-2118.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., and Leitao, P. J., 2013, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance: Ecography, v. 36, no. 1, p. 27-46.

Fenicia, F., Kavetski, D., Savenije, H. H., Clark, M. P., Schoups, G., Pfister, L., and Freer, J., 2014, Catchment properties, function, and conceptual model representation: is there a correspondence?: Hydrological Processes, v. 28, no. 4, p. 2451-2467.

Gautam, D. K., and Phaiju, A. G., 2013, Community based approach to flood early warning in West Rapti River Basin of Nepal: IDRiM Journal, v. 3, no. 1, p. 155-169.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., and Aryal, J., 2019, Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables: Fire, v. 2, no. 3, p. 50.

Guo, H., Hu, Q., Zhang, Q., and Feng, S., 2012, Effects of the three gorges dam on Yangtze river flow and river interaction with Poyang Lake, China: 2003–2008: Journal of Hydrology, v. 416, p. 19-27.

Hagen, T., 1969, Report on the Geological survey of Nepal, Vol. 1: preliminary reconnaissance: Denkschriften der Schweizerischen Naturforschenden Gesellschaft Memoires de la Societe Helvetique des Sciences Naturelles, v. 84, no. 1, p. 185.

Herder, C., 2013, Impacts of land use changes on the hydrology of Wondo Genet catchment in Ethiopia.

Hong, H., Pradhan, B., Sameen, M. I., Kalantar, B., Zhu, A., and Chen, W., 2018, Improving the accuracy of landslide susceptibility model using a novel region-partitioning approach: Landslides, v. 15, no. 4, p. 753-772.

Huang, X., Liu, J., Zhang, Z., Fang, G., and Chen, Y., 2019, Assess river embankment impact on hydrologic alterations and floodplain vegetation: Ecological Indicators, v. 97, p. 372-379.

Huang, Y., and Zhao, L., 2018, Review on landslide susceptibility mapping using support vector machines: Catena, v. 165, p. 520-529.

Joshi, N., and Dongol, R., 2018, Severity of climate induced drought and its impact on migration: a study of Ramechhap District, Nepal.

Karki, S., Koirala, M., Pradhanz, A. M. S., Thapa, S., Shrestha, A., and Bhattarai, M., 2011, GIS-based flood hazard mapping and vulnerability to climate change assessment: A case study from Kankai Watershed, Eastern Nepal: Lalitpur: Nepal climate change Knowledge Management Center Nepal Academy of Science and Technology.

Katrutsa, A., and Strijov, V., 2017, Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria: Expert Systems with Applications, v. 76, p. 1-11.

Khanal, N. R., Shrestha, M., and Ghimire, M., 2007, Preparing for flood disaster: mapping and assessing hazard in the Ratu Watershed, Nepal: International Centre for Integrated Mountain Development (ICIMOD).

Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., and Bui, D. T., 2018, A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran: Science of the Total Environment, v. 627, p. 744-755.

Kuss, M., 2006, Gaussian process models for robust regression, classification, and reinforcement learning: echnische Universität Darmstadt Darmstadt, Germany.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W., 2005, Applied linear statistical models, McGraw-Hill Irwin Boston.

Lee, S., Kim, J.-C., Jung, H.-S., Lee, M. J., and Lee, S., 2017, Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea: Geomatics, Natural Hazards and Risk, v. 8, no. 2, p. 1185-1203.

Liao, X., and Carin, L., 2009, Migratory logistic regression for learning concept drift between two data sets with application to UXO sensing: IEEE Transactions on Geoscience and Remote Sensing, v. 47, no. 5, p. 1454-1466.

Marston, R., Kleinman, J., and Miller, M., 1996, Geomorphic and forest cover controls on monsoon flooding, central Nepal Himalaya: Mountain Research and Development, p. 257-264.

Miles, J., 2014, Tolerance and variance inflation factor: Wiley StatsRef: Statistics Reference Online.

Mohammadzadeh, D., Bazaz, J. B., and Alavi, A. H., 2014, An evolutionary computational approach for formulation of compression index of fine-grained soils: Engineering Applications of Artificial Intelligence, v. 33, p. 58-68.

Msabi, M. M., and Makonyo, M., 2021, Flood susceptibility mapping using GIS and multi-criteria decision analysis: a case of Dodoma region, central Tanzania: Remote Sensing Applications: Society and Environment, v. 21, p. 100445.

Mukerji, A., Chatterjee, C., and Raghuwanshi, N. S., 2009, Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models: Journal of Hydrologic Engineering, v. 14, no. 6, p. 647-652.

Paananen, T., Piironen, J., Andersen, M. R., and Vehtari, A., Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution, *in* Proceedings The 22nd International Conference on Artificial Intelligence and Statistics2019, PMLR, p. 1743-1752.

Paquette, J., and Lowry, J., 2012, Flood hazard modelling and risk assessment in the Nadi River Basin, Fiji, using GIS and MCDA: The South Pacific Journal of Natural and Applied Sciences, v. 30, no. 1, p. 33-43.

Rahmati, O., and 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, v. 31, no. 5, p. 1473-1487.

Ren, F., Wu, X., Zhang, K., and Niu, R., 2015, Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China: Environmental Earth Sciences, v. 73, no. 8, p. 4791-4804.

Roy, P., Chakrabortty, R., Chowdhuri, I., Malik, S., Das, B., and Pal, S. C., 2020, Development of different machine learning ensemble classifier for gully erosion susceptibility in Gandheswari Watershed of West Bengal, India: Machine Learning for Intelligent Decision Science, p. 1-26.

Sachdeva, S., Bhatia, T., and Verma, A., Flood susceptibility mapping using GIS-based support vector machine and particle swarm optimization: A case study in Uttarakhand (India), *in* Proceedings 2017 8th International conference on computing, communication and networking technologies (ICCCNT)2017, IEEE, p. 1-7.

Sahana, M., and Patel, P. P., 2019, A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower Kosi River Basin in India: Environmental Earth Sciences, v. 78, no. 10, p. 1-27.

Sahana, M., and Sajjad, H., 2019, Vulnerability to storm surge flood using remote sensing and GIS techniques: A study on Sundarban Biosphere Reserve, India: Remote Sensing Applications: Society and Environment, v. 13, p. 106-120.

Samanta, R. K., Bhunia, G. S., Shit, P. K., and Pourghasemi, H. R., 2018, Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India: Modeling Earth Systems and Environment, v. 4, no. 1, p. 395-408.

Shahin, K. A., and Hassan, N., Sources of shared variability among body shape characters at marketing age in New Zealand White and Egyptian rabbit breeds, *in* Proceedings Annales de zootechnie2000, Volume 49, EDP Sciences, p. 435-445.

Sharma, K. P., Moore, B., and Vorosmarty, C. J., 2000, Anthropogenic, climatic, and hydrologic trends in the Kosi Basin, Himalaya: Climatic Change, v. 47, no. 1, p. 141-165.

Sharma, T. P. P., Zhang, J., Koju, U. A., Zhang, S., Bai, Y., and Suwal, M. K., 2019, Review of flood disaster studies in Nepal: A remote sensing perspective: International journal of disaster risk reduction, v. 34, p. 18-27.

Shekhar, S., and Pandey, A. C., 2015, Delineation of groundwater potential zone in hard rock terrain of India using remote sensing, geographical information system (GIS) and analytic hierarchy process (AHP) techniques: Geocarto International, v. 30, no. 4, p. 402-421.

Shrestha, V. P., 2007, A concise geography of Nepal, Mandala Publications.

Sihag, P., Jain, P., and Kumar, M., 2018, Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression: Modeling earth systems and environment, v. 4, no. 1, p. 61-68.

Sihag, P., Tiwari, N., and Ranjan, S., 2017, Modelling of infiltration of sandy soil using gaussian process regression: Modeling Earth Systems and Environment, v. 3, no. 3, p. 1091-1100.

Sinha, R., Bapalu, G., Singh, L., and Rath, B., 2008, Flood risk analysis in the Kosi river basin, north Bihar using multi-parametric approach of analytical hierarchy process (AHP): Journal of the Indian Society of Remote Sensing, v. 36, no. 4, p. 335-349.

Szwagrzyk, M., Kaim, D., Price, B., Wypych, A., Grabska, E., and Kozak, J., 2018, Impact of forecasted land use changes on flood risk in the Polish Carpathians: Natural Hazards, v. 94, no. 1, p. 227-240.

Tabari, H., 2020, Climate change impact on flood and extreme precipitation increases with water availability: Scientific reports, v. 10, no. 1, p. 1-10.

Tehrany, M. S., and Kumar, L., 2018, The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods: Environmental Earth Sciences, v. 77, no. 13, p. 1-24.

Tiwari, P. C., and Joshi, B., 2012, Natural and socio-economic factors affecting food security in the Himalayas: Food Security, v. 4, no. 2, p. 195-207.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M., and Savenije, H., 2015, Predicting the ungauged basin: model validation and realism assessment: Frontiers in Earth Science, v. 3, p. 62.

Vapnik, V., 1995, The nature of statistical learning theory springer new york google scholar: New York.

Wang, G., Chen, X., and Chen, W., 2020, Spatial prediction of landslide susceptibility based on gis and discriminant functions: ISPRS International Journal of Geo-Information, v. 9, no. 3, p. 144.

Wang, H., and Xu, D., 2017, Parameter selection method for support vector regression based on adaptive fusion of the mixed kernel function: Journal of Control Science and Engineering, v. 2017.

Xing, B., Gan, R., Liu, G., Liu, Z., Zhang, J., and Ren, Y., 2016, Monthly mean streamflow prediction based on bat algorithm-support vector machine: Journal of hydrologic engineering, v. 21, no. 2, p. 04015057.

Xu, M., and Li, C., 2020, Influencing factors analysis of water footprint based on the extended STIRPAT model, Application of the Water Footprint: Water Stress Analysis and Allocation, Springer, p. 105-126.

Yalçın, G., 2002, Analysing flood vulnerable areas with multicriteria evaluation.

Yang, L., and Cervone, G., 2019, Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event: Soft Computing, v. 23, no. 24, p. 13393-13408.

You, Q., Kang, S., Aguilar, E., and Yan, Y., 2008, Changes in daily climate extremes in the eastern and central Tibetan Plateau during 1961–2005: Journal of Geophysical Research: Atmospheres, v. 113, no. D7.

Youssef, A. M., Pradhan, B., Jebur, M. N., and El-Harbi, H. M., 2015, Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia: Environmental Earth Sciences, v. 73, no. 7, p. 3745-3761.

Zhao, K., Popescu, S., Meng, X., Pang, Y., and Agca, M., 2011, Characterizing forest canopy structure with lidar composite metrics and machine learning: Remote Sensing of Environment, v. 115, no. 8, p. 1978-1996.