Journal of Information Systems Engineering and Management
2025, 10(14s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Effectiveness of Random Forest Model for Flash Flood Susceptibility in the Himalayan Region

Hemwati Nandan Bahuguna[1*], Sushma Bahuguna[2], Syed Hameedur Rahman Zaini[3], Karina Bhatia Kakkar[4]

[1] *Doctoral Scholar, School of Management, GD Goenka University, Gurgaon, India*

[2] *Professor, Department of Information Communication and Technology, TIAS, Aff. to GGSIP University, Delhi, India*

[3] *Asst. Professor, School of Management, GD Goenka University, Gurgaon, India*

[4] *Associate Professor, School of Management, GD Goenka University, Gurgaon, India*

*** Corresponding Author: h.n.bahuguna11@gmail.com**

ARTICLEINFO

Received: 22 Nov 2024

Revised: 08 Jan 2025

Accepted: 22 Jan 2025

ABSTRACT

Machine learning (ML) approach is being increasingly recognized as vital tool for disaster risk reduction owing to its ability to address both the scale and the impact of a disaster. Risk of flash floods is currently a major problem across the Himalayan region. In present study we evaluated effectiveness of Random Forest (RF) model for flash flood susceptibility, based on real-world data in the Indian Central Himalaya Region (ICHR), where recurrent flash floods are being experienced every year. A geospatial dataset including 200 flash flood locations and eight conditioning factors namely elevation, slope, aspect, profile curvature, distance from river, annual rainfall, land use land cover (LULC) and lithology was used for performance evaluation of Random Forest model for flash flood susceptibility assessment. The effectiveness of the model is evaluated using area under the ROC curve (0.922), accuracy (0.925), precision (0.903), recall (0.921) and F-score (0.911) metrics. The results show that random forest could be an effective tool for flash flood susceptibility assessment in the Indian Central Himalayan Region. Furthermore, by considering optimum conditioning factors based on topographical, geological and hydrological conditions, the model can be used by managers and planners for flash flood management and sustainable conservation of the human society in the other Himalayan regions.

Keywords: Ensemble machine learning, Geospatial dataset, Conditioning Factors, Mountainous Region

1. INTRODUCTION

Flash flooding, a specific type of flooding that occurs when large amount of water is discharged within a few minutes or hours (three to six hours) of excessive rain fall, the collapse of natural ice or dam failure (Wang et al., 2019) is responsible for more than 5000 deaths annually on global basis (Modrick & Georgakakos 2015). Flash floods strike Himalayan Region, predominantly during monsoon period resulting in huge loss of life, property, environment, infrastructure, agriculture land and other facilities (NIDM, 2015). One of the major flash floods incidences of the region on 16-17 June 2013 resulted in a major tragedy in which human death toll surpassed 4000 and was a major setback for the economy of the region (Rautela, 2016). Flash floods of the region are often associated with cloudbursts, landslide dam outbursts and glacial lake outbursts. Flash flood susceptibility assessment is essential for mitigation measures and risk management strategies to reduce future losses (Youssef et al., 2016). Flood susceptibility is qualitative or quantitative assessment delineating spatial distribution of the probability of likely flood events in the region (Rehaman et al., 2019). It measures the likelihood of future flood events depending on the meteorological conditions (Quinn et al., 2011).

With continuous development of geographic information system (GIS) and machine learning (ML) approach, researchers have been applying different Machine Learning models for flash flood susceptibility in mountainous regions of the world. Also, few studies have been conducted by researchers in the ICHR using ML techniques. Singh and Pandey (2022) applied weighted sum approach (WSA), principal component analysis (PCA) and integrated approach (IA) for morphometric characterization and subsequent sub-watersheds vulnerability zonation in the upper Ganga catchment area of ICHR. Using hydrologic and demographic data, Vishwanath and Tomaszewski (2018) have employed analytical hierarchy process for flash flood hazard, vulnerability and risk assessment in the region. Rana and Mahanta (2022) have applied ensemble approaches to bivariate statistical model and machine learning techniques such as artificial neural network, support vector machine and K- nearest neighbor models for flash flood susceptibility modeling in the region. Sachdeva et al. (2017) used SVM and particle swarm optimization ML techniques for flood susceptibility assessment in Chamoli district of Uttarakhand. In present study effectiveness of ensemble ML algorithm Random Forest is assessed for flash flood susceptibility assessment in the ICHR. Random Forest algorithm is chosen for modeling as it is a versatile and powerful machine learning algorithm with a number of advantages; ability to learn non-linear decision boundary, high accuracy, flexibility and robustness, feature importance, Scalability and parallel processing.

2. STUDY AREA

Indian Central Himalayan Region located between 28° 44′- 31° 28′ N latitude and 77° 35'- 81° 01'E longitudes, includes 320 km long stretch of mountains between Tons-Pabbar valleys forming the eastern border of the Himachal Pradesh in the west and Kali River forming Indo-Nepal border in the east (Figure 1). It covers 53,483 Sq. km area of Uttarakhand state and is also known as Uttarakhand Himalaya. The region having 93% mountainous region comprises of two distinct geographical entities namely Garhwal and Kumaon Himalayas with 13 districts, 78 tehsils and 95 community development blocks. The population of the region is 101 million as per 2011 census. The region shares boundary with Indian state of Himanchal Pradesh in NW, Utter Pradesh in SW and a short boundary with Haryana in west. It is bordered by autonomous region of Tibet (China) in North and by Nepal in East. Topographically the region is characterized by glaciers, large number of perennial rivers, rugged mountainous terrain and dense forests. The region is birth place of several perennial rivers and is dominated by three river systems; The Ganga System, Yamuna-Tons River System and Kali River System. Tectonically the region is delineated by Main Central Thrust (MCT), Main Boundary Fault (MBT) and Main Frontal Thrust (MFT) dividing it into Great Himalaya, Lesser Himalaya and Outer Himalaya. The elevation of the region varies between 300 to 8000 meters. The region is covered by evergreen forests (68.4%), and mountains (85%) (Bhambri et al., 2016). The region is conferred with relatively average annual rainfall of 1229 mm^2 and average temperature of the region varies from 1.7^{o}C to 42^{o}C (NIDM, 2015).

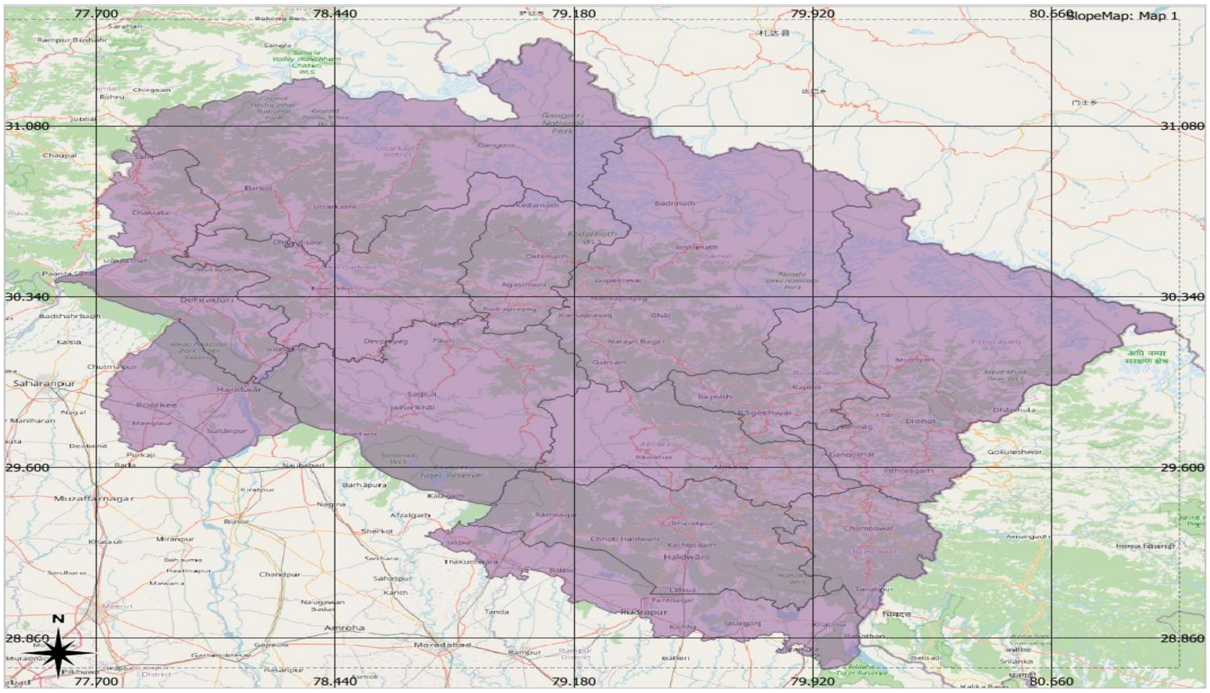


Figure 1. Location map of study area

3. METHODOLOGY

Methodology applied to evaluate effectiveness of Random Forest model for flash flood susceptibility includes three main steps (Figure 2): (i) construction of geospatial dataset for flash flood locations and conditioning factors; (ii) development of Random Forest model and (iii) evaluation of model performance. A detailed description of the steps is presented in in the following subsections.

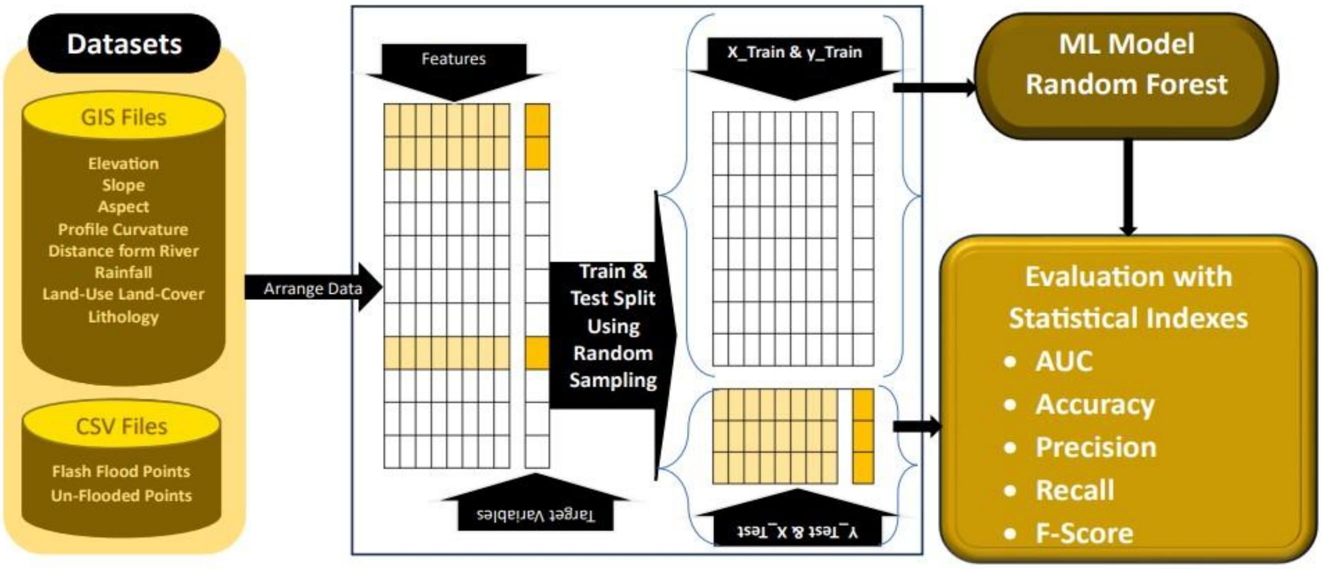


Figure 2: Flow chart of modeling methodology

3.1 GEOSPATIAL DATASET

3.1.1 Flash Flood Inventory

In present research 200 recent and historical flash flood locations were obtained based on extensive review of literature. Table 1 illustrates the list of reported flash floods from 1983 to 2023 in the Indian Central Himalayan Region (NIDM, 2015; Rautela, 2016; Joshi and Kumar, 2006; Sati, 2007; Prakash, 2015; Dimri et al., 2017; Kumar et al., 2018; Sati et al., 2020; Khanduri, 2020, 2021, 2022; Singh et al., 2022; Sati and Kumar, 2022; SANDRP). Similar to the flood locations, 200 of unflooded locations were chosen randomly for the study.

	Location, District	**Year**	**Impacts**
1	Birahi, Chamoli	1893, 6 September	Birahi village was washed away
2	Satpuli, Pauri	1951, 14 September	Swept away convoy of 22 Buses and trucks along with 30 drivers and conductors
3	Belakuchi, Chamoli	1970, 20 July	Belakuchi settlement washed off along with convoy of 36 buses, 13 bridges swept away and lower portion of Srinagar town was destroyed
4	Dobata , Pithoragarh	1971, 19 July	12 killed, 37 buildings damaged
5	Joshiyara, Uttarakashi	1978, 05 August	Shops and houses in Joshiara village swept away, settlements and agriculture land washed away, heavy damage to Maneri-Bhali HEP (Kandoliya Gad Flash flood)
6	Saikot, Chamoli	1979, 17 June	3 persons killed,70 cattle lost and 20 houses damaged
7	Kuntha, Rudrprayag	1979, 17 August	39 dead, 39 cattle lost and 20 houses damaged

8	Sirwari,Rudraprayag		13 persons killed,150 cattle lost and 34 houses damaged
9	Gyansu Nala, Uttarakashi	1980, 24-25 June	24 persons killed and washed away many houses
10	Mandakhal, Pauri	1982, 31 July	3 persons killed, 80 animals lost and 8 houses damaged
11	Karmi, Almora	1983, 22 July	37 dead, several houses washed away
12	Kapkot, Bageshwaer	1983, 23 July	25 dead, 20 animals lost and damaged 6 houses
13	Karnprayag, Chamoli	1989, 05 Sept	3 dead
14	Neelkanth,Pauri	1990, 09 July	Killed 100 people and destroyed 10 houses
15	Dewar Khandora, Chamoli	1991, 16 August	26 dead, loss of 63 animals and 38 houses damaged
16	Khankrakhat hill, Chamoli	1992, 02 September	Gadini Bazar collapsed, 14 people dead, 18 shops and 13 water mills washed away
17	Patidy Hill, Chamoli	1993, 22 July	Heavy loss of animals and property
18	Chaukhuti, Almora	1994, 30 July	4 people dead
19	Bhintai, Pauri	1995, 13 August	Killed 13 people and damaged 6 houses
20	Berinaga, Pithoragarh	1996, 17 July	18 people killed and 85 houses damaged
21	Raitoli,Pithoragarh	1996, 26 July	Killed 16 person
22	Neelkanth,Pauri	1997, 02 August	18 people dead
23	Malpa, Pithoragarh	1998, 17 August	Killed 221 people including 60 Kailash Mansarovar pilgrims
24	Paundar, Rudraprayag	1998, 19 August	Killed 103 people, 422 animals lost and 820 houses damaged
25	Pujargaon, Uttarakashi	2000, July	Road and agriculture fields damaged
26	Phata, Rudraprayag	2001, 16 July	27 people killed, 64 animals lost and 22 houses damaged
27	Byung, Rudraprayag		
28	Gona, Tehri	2001, 31 August	7 persons killed, 12 injured,7 cattle lost and 28 houses damaged
29	Khet, Pithoragarh	2002, 12 July	4 persons killed, 1 missing, 20 animals lost and 7 houses damaged
30	Budhakedar, Tehri	2002, 10 August	Lower part of Budha Kedar washed away along with agriculture fields
31	Agunda, Tehri		28 people killed, 99 cattle lost, 151 houses and micro-hydro power plant damaged, more than 17 villages and 1200 people were impacted by this event
32	Marwari, Tehri		
33	Thati, Tehri		
34	Bhatwari, Uttarakashi	2002, 19-20 August	Killed 5 people and 26 animals lost
35	Dunda, Uttarakashi		
36	Didihat,pithoragarh	2002,20 July	4 people dead
37	Sarnol, Uttarakashi	2003, 29 August	207 animals lost

38	Kapkot, Bageshwar	2004, 21 may	3 people dead
39	Kapkot, Bageshwar	2004, 9 June	3 people dead
40	Badrinath, Chamoli	2004, 24 September	A stretch of 200 m of NH-58 was completely destroyed.
41	Teela, Pauri	2005, 21 June	Killed 30 animals and destroyed 6 houses
42	Govindghat, Chamoli	2005, 29-30 June	11 people dead, more than 200 vehicles,20 shops, hotels and human settlement washed away
43	Kedarnath, Rudraprayag	2005, 05 July	10 shops destroyed and 4 shops damaged partially
44	Jangal Chatti, Rudraprayag	2005, 07 July	1 person killed and 1 house damaged
45	Vijaynagar, Rudraprayag	2005, 21 July	4 dead and 14 houses destroyed
46	Jhuni, Bageshwar		1 person killed, 3 cattle lost and 6 houses destroyed
47	Rathi, Pithoragarh	2005, 18 August	8 persons killed and 129 animals lost
48	Atholi, Uttarakashi	2005, 27 August	15 animals lost and 3 cow sheds destroyed
49	Devali, rudraprayag	2006, 26 July	19 animals killed, 8 houses and 6 shops damaged
50	Ladoli, Rudraprayag		
51	Gholtir, Rudraprayag		
52	Chamoli Tehsil, Chamoli	2006, 01 August	18 houses damaged
53	Devpuri, chamoli	2007, 12 July	8 persons reported buried under debris
54	Didihat,pithoragarh	2007, 13 August	4 people dead
55	Baram, Pithoragarh	2007, 06 September	5 persons were killed and 9 others feared dead
56	Joshimath, Chamoli	2009, 25 July	3 people dead
57	Kuity, Pithoragarh	2009,18 August	Wiped out 2 villages (Jhakhla and Leh) claiming 43 lives
58	Kot, Pauri	2010, 19, July	6 people dead
59	Dhari, Nainital	2010, 18 August	13 people dead and 3 injured
60	Dhari, Nainital	2010,20 September	5 people dead and 12 injured
61	Kot, Pauri	2010, 22 September	5 people dead
62	Nayalgarh, Pauri	2010, 11 September	3 people dead
63	Karnprayag, Chamoli	2010, 08 September	3 people dead

64	Sumgarh, Bageshwar	2010, 18 September	18 children of primary school buried under rock debris
65	Tuneda, Bageshwar	2011, 18, August	21 people dead
66	Gauri Kund, Rudraprayag	2012, 04 July	Road damaged
67	Chaumasi, Rudraprayag	2012, 19 July	1 house damaged
68	Pandrasu Ridge, Uttarakashi	2012, 3 August	35 human lives lost, 20 people injured and 436 livestock dead, over 787-hectare agriculture land washed away
69	Sangam Chatti, Uttarakashi		
70	Bhatwari, Uttarakashi		
71	Okhimath, Rudraprayag	2012, 13-14 September	69 human lives lost, 15 injured,70 houses destroyed
72	Kimana, Rudraprayag		
73	Mangoli, Rudraprayag		
74	Chunni, Rudraprayag		
75	Phata, Rudraprayag	2013, 16-17 June	Majoy tragedy surpassed death toll 4000. Rudraprayag, Chamoli, Uttarakashi, Pithoragarh and Bageshwar were worst effected districts of the state. There was heavy rainfall in the entire state with the onset of early monsoon in June 2013. Heavy rainfall between 14th and 18th June 2013 resulted in flash flood and landslides at multiple locations that turned into massive disaster in the state. There was severe devastation particularly in the Mandakini valley in Kedarnath-Rambara-Gauri kund area due to breach of Chorabari lake. The incidence booked massive losses of human lives, infrastructure and property (*100 persons dead, 4019 persons missing, 183 persons injured, 10292 Farm animal lost, 1985 fully damaged houses, 2018 severely damaged houses, 5590 partially damaged houses, 741 hectares silted agriculture land, 10696 hectares agriculture land lost and 497 hectares crop loss. The disaster was a major setback for the economy of the state*).
76	Mangoli, Rudraprayag		
77	Kimama, Rudraprayag		
78	Kedarnath, Rudraprayag		
79	Rambara, Rudraprayag	2013, 16-17 June	
80	Gauri Kund, Rudraprayag		
81	Sonprayag, Rudraprayag		
82	Lambagad, Chamoli		
83	Govindghat, Chamoli		
84	Pulna, Chamoli		
85	Tharali, Chamoli		
86	Narayabagad, Chamoli		
87	Dharchula, Pithoragarh		
88	Munsyari, Pithoragarh		
89	Balwakot, Pithoragarh		
90	Madkot, Pithoragarh		
91	Pangla, Pithoragarh		
92	Tawaghat, Pithoragarh		
93	Baram, Pithoragarh		

94	Harshil, Uttarkashi		
95	Sonprayag, Rudraprayag	2015, 27 June	1 Bridge washed away
96	Jaimandi, Rudraprayag		Houses, Bridges, roads and agriculture land washed away
97	Kothiyara, Tehri	2016, 28 May	100 animals buried and 50 residential houses brought under load of debris
98	Kemra, Tehri		20 houses destroyed
99	Silyara, Tehri		50 houses were brought under debris, motor bridge washed away
100	Kumalgaon, Pithoragarh	2016, 01 July	3 persons killed and 70 animals killed
101	Bastari, Pithoragarh		19 persons killed, 174 cattle lost and 16 buildings damaged
102	Tatalgaon, Almora	2017, 26 May	1 house damaged and 8 domestic animals dead
103	Bijrani, Almora		
104	Mangti, Pithoragarh	2017, 14 August	9 lives lost, 18 persons missing
105	Malpa, Pithoragarh		
106	Mona Chida, Chamoli	2018, 02 May	Several vehicles damaged; debris rubble damaged several houses
107	Paithani, Pauri	2018, 01 June	Damaged cow shed killing 4 cattle
108	Betalghat, Nainital		Rubble and muck dumped inside homes and over fields of the Katmi and Gjar villages
109	Jauljibi, Pithoragarh		Many houses and shops damaged
110	Balati, pithoragarh	2018, 02 July	Seraghat hydro power project was damaged
111	Seemadwar, Dehradun	2018, 11 July	7 people dead 2 injured and 2 houses collapsed
112	Tharali, Chamoli	2018, 16 July	15 houses, 10 vehicles damaged, 2 ropeways and 1 road bridge washed away, several mini hydro projects affected
113	Yamnotri, Uttarakashi	2018, 17 July	The foot bridge connecting to the shrine was washed away, Kali Kamli Dharamshala was severely damaged and hot water Kunds were filled with debris
114	Malari, Chamoli	2018, 19 July	2 persons dead, 5 trapped in debris, over 150 m long Joshimath- Malari Road stretch washed away
115	Lambagad, Chamoli	2019, 02 June	Killing an 82 years old shepherd after flash floods in Gangani stream, damages to agricultural lands were also reported.
116	Khira, Almora		A person was missing and some cattle also washed away apart from large scale destruction.
117	Mori, Uttarakashi	2019, 21 June	One person was reported dead and four other injured due to the incident, flood water also entered in some of

			the houses and shops apart from damaging trees and blocking the main road for a couple of hours.
118	Mansari, Pauri	2019, 23 June	Swept away cattle and damaging homes
119	Sari, Rudraprayag	2019, 04 July	Farm lands, 40-meter part of main road and water supply pipe line damaged
120	Gairsain, Chamoli	2019, 06 August	Damaged bridge, school building and cow sheds
121	Padmalla, Chamoli	2019, 08-09 August	Killed a women and a kid, damaged 12 houses, 2 dozen cow sheds, 6 foot bridges, 10 water mills along with agriculture lands in Padmalla, Faldiya, Ulangra,Tallore and Bamanbera villages
122	Saikot, Chamoli		2 dead, 5 severely injured, damaged houses and cow sheds
123	Tharthi, Tehri		
124	Aali, Chamoli	2019, 12 August	1 dead and several houses damaged
125	Lankhi, Chamoli		3 persons were killed by landslide after flash flood
126	Hawil Kulwan, Bageshwar		Several houses and cow sheds washed away, destroyed farm lands and drinking water supply
127	Arakot, Uttarakashi	2019, 18 August	21 people killed and 74 animals lost (Arakot, Sanel,Makudi, Tikochi villages), 2 motor bridges damaged
128	Timtia, Pithoragarh	2019, 06-07 September	A man was killed and 3 others injured
129	Goind Ghat, Chamoli		Several vehicles were buried under debris while 30 m stretch of NH 58 completely damaged
130	Gudam, Chamoli		Houses and cattle sheds destroyed, several acre of farm lands was destroyed
131	Dhurma, Chamoli	2019, 07-08 September	A house and 2 water mills washed away, 6 houses and intercollege building damaged
132	Clement Town, Dehradun	2019, 27 September	About 15 houses damaged, rivulets and drains swelled dangerously
133	Dharchula, Pithoragarh	2020, 14 July	Huge landslides and damage to road of India-China border, landslide blocked river forming dam
134	Madkot, Pithoragarh	2020,18-19 July	3 killed, 6 injured, 11 people washed away (Madok, Tanga), highway and border roads left blocked. 3 houses buried in Gaila village, Bata and Sirtaul villages were also hit by the incident damaging 5 houses and killing of cattle.
135	Tanga, Pithoragarh	2020, 19 July	11 people killed, losses of property
136	Patherkot, Pithoragarh		3 persons killed, massive property losses
137	Bangapani, Pithoragarh	2020, 28 July	Damaged homes, farmlands, local roads and bridges
138	Pader, Chamoli		A women buried, huge property losses
139	Boora, Chamoli		Debris entered Boora village, damaged houses and cowsheds

140	Sirwadi, Rudraprayag	2020, 09 August	7 houses and large area of agriculture land was affected, Village roads, irrigation channel, drinking water supply, electric poles washed away.
141	Gangi, Tehri		20 cattle were buried in the debris and rubble
142	Aare, Bageshwar	2020, 10 August	Home damaged and cowshed buried under the muck, damages to local roads and bridges
143	Near Lakhwar Dam, Dehradun	2020, 19 August	Bridge damaged
144	Tali Kansari, Chamoli	2020, 24 August	House damaged, 1 died 5 injured
145	Raini,Chamoli	2021, 07 February	83 dead and 121 missing (workers of Tapovan – Vishnughat HEP site)
146	Tapovan -Vishnughat HEP,Chamoli		
147	Devprayag, Tehri	2021, 11 May	Shops and houses crumble into swollen river
148	Dovalya, Rudraprayag	2021, 01 June	Agriculture land and one bridge destroyed
149	Maldeota, Dehradun		The resultant deluge flooded the Maldevta junction area with massive amount of debris and sludge. The muck also entered several houses, hotels, restaurants in the area and blocked Dwara, PCL roads.
150	Sauni Binsar, Almora		Damage to boundary wall, campus and gate of Swargasharm temple
151	Narkota, Rudraprayag	2021,12 June	3 killed, 3 injured and 4 missing, Six homes some cowsheds and farmlands damaged.
152	Bin, Pithoragarh		Damage to homes, farm crops, drinking water supply lines and local road
153	Parsari, Chamoli	2021, 23 JUNE	Washed away agriculture land, dumped land rocks and debris on the Niti-Joshimath road
154	Dharchula, Pithoragarh	2021,08 July	Tanakpur-Tawaghat highway got washed away.
155	Basti, Bageshwar	2021,17 July	Crops on several farmlands belonging to a dozen families of Sangad Basti village. Roads and irrigation channels in the village area were also affected in the incident.
156	Nirakot, Uttarakashi	2021, 18-19 July	Motor bridge, foot bridge and cow shed washed away
157	Mando,Uttarakashi		Affected 30 families in Mando village and damaged around 14 houses in Kankradi village
158	Gangori, Uttarakashi		Dozen homes were flooded with muck, Damage to agriculture land and crop
159	Maid, Tehri	2021, 19 July	Injured 1 and buried portions of seven houses under muck and rubble, affected about 30 naali agricultural land and damaged a foot bridge.
160	Valley of Flowers, Chamoli	2021, 08 August	Washed away 20-meter pathways and a foot bridge in Bamandhon near glacier point.

161	Marchula, Almora	2021, 13 August	The incident damaged roads and impacted homes, shops, courtyards, village pathways, drinking water pipelines
162	Sarkhet, Dehradun	2021, 24 August	More than 40 homes were flooded
163	Timla, Tehri		10 houses destroyed, damage to agriculture land and bridge
164	Jakhan, Dehradun	2021, 26 August	A home collapsed killing an elderly villager. Several farmlands of village were filled with silt, sludge destroying the crop.
165	Binhar,Dehradun	2021, 26-27 August	Impacted under construction Vyasi Hydro Power Project
166	Bhitareli, Dehradun	2021, 27 August	Two cow sheds washed away and caused landslides at half a dozen places in the area. Heavy rainfall, flash flood spells left behind huge trail of destruction impacting local roads, bridges and agricultural land and crops.
167	Jumma, Pithoragarh	2021, 30 August	2 people killed and 5 buried in Jumma Village
168	Syunsai, Pauri	2021, 07 September	The deluge that rolled down the hill damaged farm lands and destroyed crops of over 15 nalis.
169	Narayabagad, Chamoli		The incident caused landslide and flood damaging several homes, shops and cow sheds with debris in the area.
170	Sirabagad, Rudraprayag	2021, 09 September	Two people went missing after a diesel truck which was hit by deluge fell into Alaknanda. 3 vehicles trapped in muck sludge.
171	Panti, Chamoli	2021, 20 September	1 injured, damaged 11 huts belonging to migrant workers of Border Roads Organization. .The muck sludge trapped several vehicles.
172	Bohrakun, Nainital	2022, 07 May	Damaged cow shed and killed a livestock. The campus of Mallikaarjun School was filled with debris
173	Papoli, Bageshwar	2022, 10 May	Muck and debris invaded half a dozen homes and farming land of several farmers.
174	Mori, Uttarakashi	2022, 05 July	Mori Market damaged
175	Rastadi, Uttarakashi		Damaged vegetable crops and part of Syori road, washed away two public toilets, flash flood debris also entered some homes.
176	Kedarkantha		Deluge in Fafrala Khad washed away about 500 metres road completely, over two dozen villages were disconnected from road services
177	Munar, Bageshwar	2022, 08 July	3 Foot bridges, 2 water mills, 2 public toilets and 1 hand pump washed away. Several houses, farmlands and vehicles damaged
178	Sain Rathi (Mynsyari), Pithoragarh		1 bridge damaged

179	Haldwani, Nainital	2022, 09 July	Severe urban flooding
180	Luhari, Dehradun	2022, 14 July	3 cattle killed
181	Valley of Flowers, Chamoli	2022, 20 July	About 163 tourists stranded.
182	Sobla, Pithoragarh	2022, 30 July	Bailey bridge of Border Roads Organization washed away
183	Jogdi, pauri	2022, 01 August	70 to 80 agriculture fields washed away
184	Basti, Rudraprayag	2022, 09 August	3 cattle killed
185	Agarchatti,Chamoli	2022, 10 August	3 houses destroyed and 10 houses damaged
186	Purola, Uttarakashi	2022, 10-11 August	8 shops washed away
187	Kalsi, Dehradun	2022, 11 August	Structures built close to stream including school damaged severely
188	Sarkhat, Dehradun	2022, 20 August	6 dead and 28 cattle washed away
189	Malyakot, Tehri		9 persons and 44 cattle dead
190	Gwar, Tehri		2 dead, 4 missing
191	Binak, Pauri		1 person, 13 cattle filled and heavy damage in Tall and Havel valleys
192	Arakot, Uttarakashi		4 houses damaged; agriculture land washed away
193	Chirbatiya, Tehri	2022, 24 August	Thatri village was flooded by Road muck and debris
194	Dharchula-Khotila, Pithoragarh	2022, 10 September	1 dead, destruction in Khotila village,
195	Sahiya, Dehradun	2022,25 September	7 houses and one bridge damaged
196	Kain khola, Pithoragarh	2022, 09 October	Agriculture land and crops damaged
197	Chal, Pithoragarh	2023,07 July	Bridge washed away and many houses damaged
198	Purola, Uttarakashi	2023, 22 July	Over 2 hectares of land and link roads washed away, 4-foot bridges collapsed
199	Gaurikund,Rudraprayag	2023, 03 August	4 people dead, 15 missing, swept away 3 shops
200	Kathgodam, Haldwani	2023,09 August	2 houses collapsed, 200 people relocated

Table1: List of reported flash floods in the Indian Central Himalayan Region

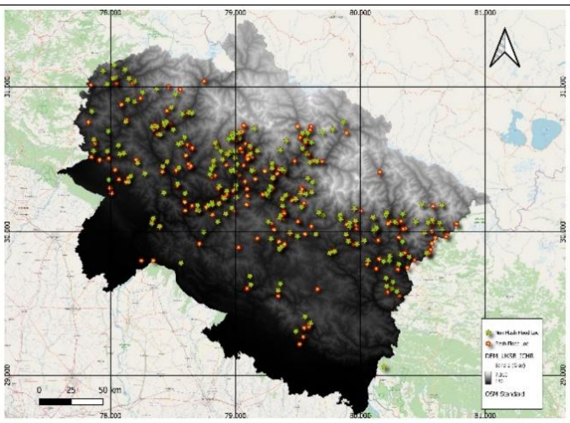
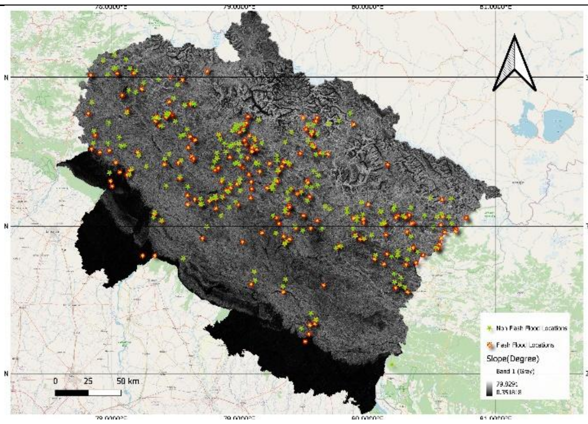
3.1.2 Flash flood Conditioning Factors

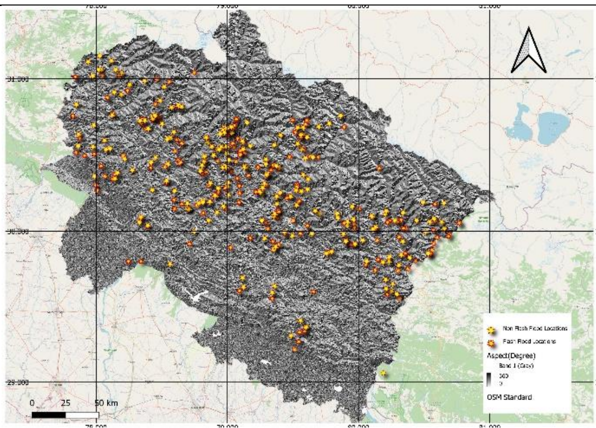
For flash flood susceptibility modeling, it is crucial to select appropriate conditioning factors. In present study a total 8 conditioning factors were selected based on nature of flash flood observation related to topographic, hydrologic and anthropogenic activity. The conditioning factors include elevation, slope, aspect, profile curvature, distance from river, LULC, annual rainfall and lithology. The digital elevation model (DEM) of study area was extracted from a 30m resolution of Shuttle Radar Topographic Mission (SRTM) data. Then dataset of conditioning factors such as elevation, slope, aspect, profile curvature was extracted using QGIS 3.34 Prizren software spatial analysis tool.

Elevation is a conditioning factor owing to the weathering of rocks and soil on the slope (Nguyen, et al. 2019). Generally, elevation and flooding have inverse relationship and low elevated regions are more susceptible to flooding (Chen et al., 2019). Elevation also acts as an indirect informer for the distribution of land cover, climate and the effects of rock weathering at various elevations. The elevation of study area ranges between 186 m to 7310 m (Figure 3a). Slope controls the speed of water flow from high to low altitude and is essential conditioning factor for studying flash flood susceptibility (Pham et al., 2019). The slope of the study area ranges from zero degree to 80 degrees (3b). Aspect pertains to the direction of water flow affecting occurrence of flash flood (Aryal et al., 2003). It is calculated in the direction of clockwise in degrees from 0 to 360, due north (Figure 3c). Profile curvature is another important factor that delineate surface with accelerated surface runoff (Abedi et al., 2022). Positive values imply a slower water flow across the surface, whilst negative values suggest a faster flow (Figure 3d).

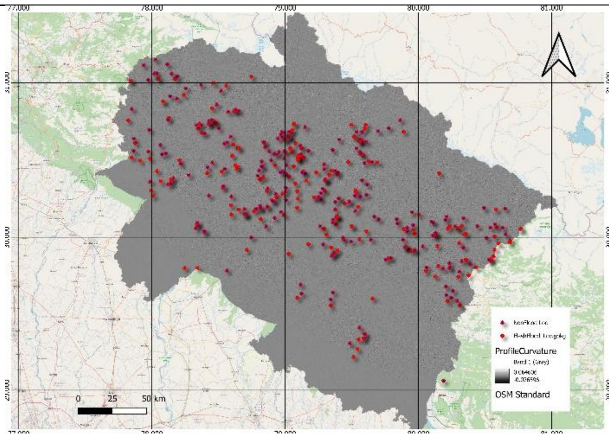
Rainfall is primary cause of floods and flood intensity increase with the increase in rainfall (Bordbar et al., 2022). Long term rainfall data (2001 to 2020) from Data Portal of Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine were used to generate annual rainfall map (Figure 3e). Lithology represents rock types that effect process of runoff and infiltration, thus affecting flash flood occurrence (Rahmati et al., 2016). Geological map of the area was constructed from world geologic map - USGS with twelve classes (Figure 3f): undivided precambrian rocks (Pc), tertiary igneous rocks (Ti), paleozoic and mesozoic metamorphic rocks (MzPz), quaternary sediments (Q), neogene sedimentary rocks (N), triassic metamorphic and sedimentary rocks (TRMS), Jurassic metamorphic and sedimentary rocks (JMS), Mesozoic intrusive (Mi), cretaceous sedimentary rocks (Ks), undiffentiated paleozoic rocks (Pz), lower Paleozoic rocks (Pzl), cretaceous and tertiary igneous and metamorphic rocks (TKIM).

Fluvial flooding has direct relationship with the distance from river and stream (Rusk et al., 2022; Vojtek & Voitekova, 2019). Flooding occurs when water level in a river or stream rises and overflows onto the neighboring land. Figure 3g depicts distance from river map for flash flood points and non- flood points of the study area. Land use Land Cover (LULC) plays significant role in hydrological and geomorphological process by influencing runoff generation, evapotranspiration, infiltration, and sediment dynamics (Maestre & Cortina, 2002; Pham et al., 2020). LULC map of study area generated from sentinel-2 is categorized into forest, grass, agriculture land, scrub/shrub, built up areas, barren land, snow/ice and water (Figure 3h).

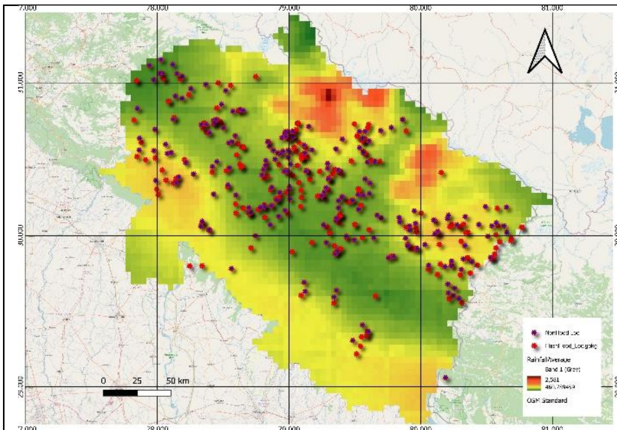
	
(a) elevation	(b) slope



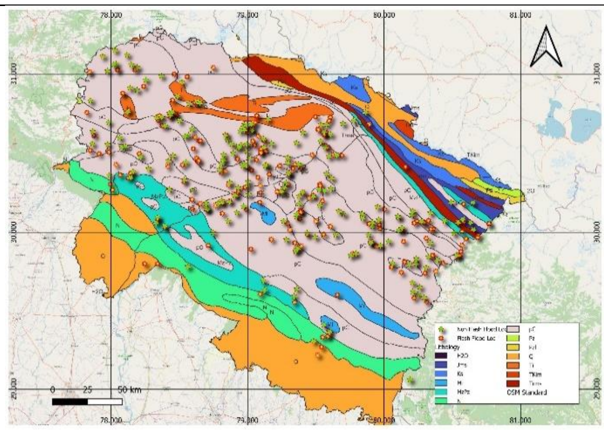
(c) aspect



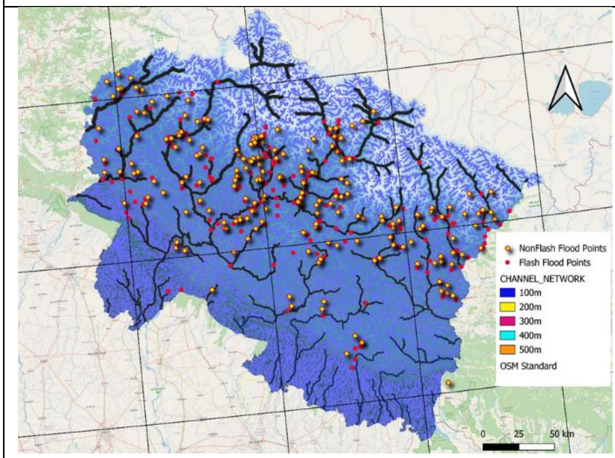
(d) Profile curvature



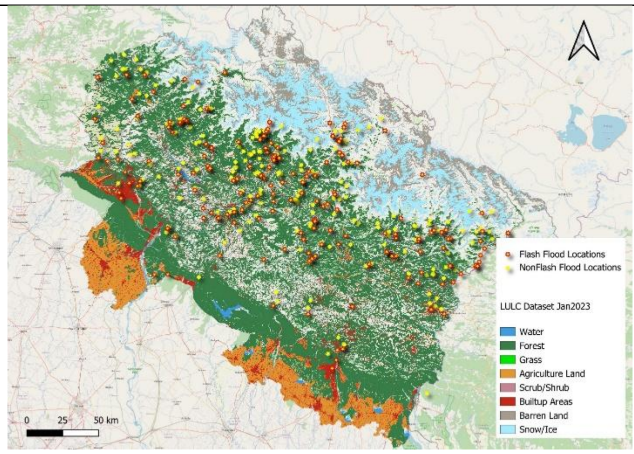
(e)Rainfall



(f) Lithology



(g) Distance from River



(h) LULC

Figure 3. Maps of flash flood conditioning factors

3.1.3 Correlation Analysis and Multicollinearity Test

The issue of multi-collinearity can significantly affect the prediction accuracy of machine learning model. To avoid this effect, correlation coefficient and Variance Inflation Factor (VIF) were calculated using Seaborn library. The output of absolute values for correlation coefficient and VIF are less than 0.5 and 2 respectively (Figure 4 and Table 2) which indicate that there is no significant relationship between the conditioning factors and can be used as input variables for flash flood susceptibility.

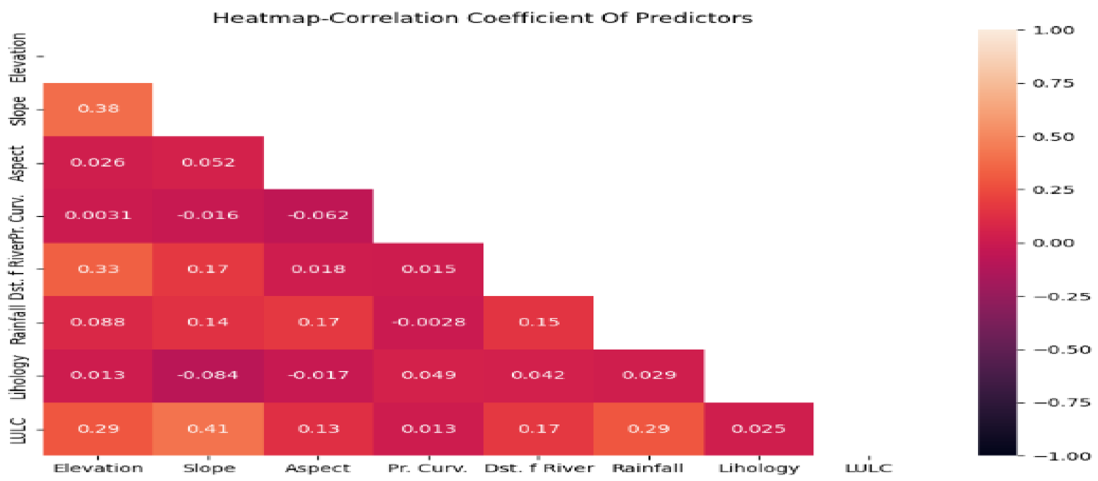


Figure 4: Correlation Coefficient matrix diagram of conditioning factors

Variable	VIF
Elevation	1.306387
Slope	1.343420
Aspect	1.041639
Profile Curvature	1.007146
Distance to River	1.145816
Rainfall	1.129478
LULC	1.340298
Lithology	1.017697

Table 2: The results of Multicollinearity Test

3.1.4 Factor Importance

The mean decrease Gini Index of RF was used to determine the relative importance of conditioning factors which is crucial for understanding the contribution of each factor and hazard risk pattern (Table 3). The results revealed that rainfall and distance from river are the most powerful factors to predict flash flood risk, while lithology and plan curvature seemed to have the least importance for flash flood susceptibility modeling.

Conditioning Factors	Relative Importance
Rainfall	0.28
Distance to river	0.24
LULC	0.12
Elevation	0.11
Slope	0.10
Aspect	0.06
Plan Curvature	0.05
Lithology	0.01

Table 3: The relative importance of conditioning factors

3.2 DEVELOPMENT OF RANDOM FOREST MODEL

Random forest (RF) is a type of supervised ensemble learning method developed by Breimen (2001) for classification and prediction. It is basically a statistically-based approach and can easily handle a large number of variables (Goetz et al., 2015). The model is used for analysis of dynamic trends known to non-linear interactions between explanatory and response variables and does not require any kind of assumption to establish relationship among explanatory and response variables (Band et al. 2020). The RF modeling classifier consist of several decision trees and has proved its high accuracy and superiority (Fawagreh et al., 2014). In order to create a group of decision trees with controlled variation, RF modeling technique combines bagging sampling methodology of Breiman (1996a) and random selection of features, introduced independently by Ho (1995) and Amit and Geman, (1997). Each decision tree in the ensemble is generated using a sample with replacement from the training data applying bagging technique and performs as a base classifier to define the class label of an unlabeled instance which is classified on the basis of majority of votes. Figure 5 illustrates a single decision tree of random forest classifier applied for training data set of the study area. In this study 70% dataset of 400 locations (200 flood points and 200 unflooded points) was used to train (training set) the model and 30% of the dataset was used to evaluate (validation dataset) the performance of the model.

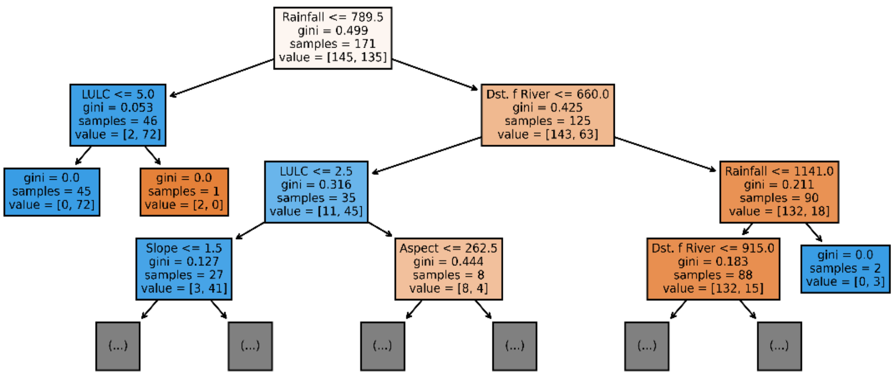


Figure 5: Illustration of single decision tree of random forest classifier applied for training dataset of study area

3.3 EVALUATION OF MODEL PERFORMANCE AND RESULTS

Area under the ROC curve (AUC) is most commonly used metric to evaluate the performance of the model for natural hazard susceptibility mapping (Abedini et al., 2019). ROC curve (Receiver operating characteristic curve) is graphical presentation of a classification model at all classification thresholds. The curve plots two parameters; True Positive Rate (TPR= TP/TP+FN) and False Positive Rate (FP=FP/FP+TN); where TP=True Positive, FP=False Positive, TN=True Negative and FN=False Negative. AUC provides an aggregate measure of all possible

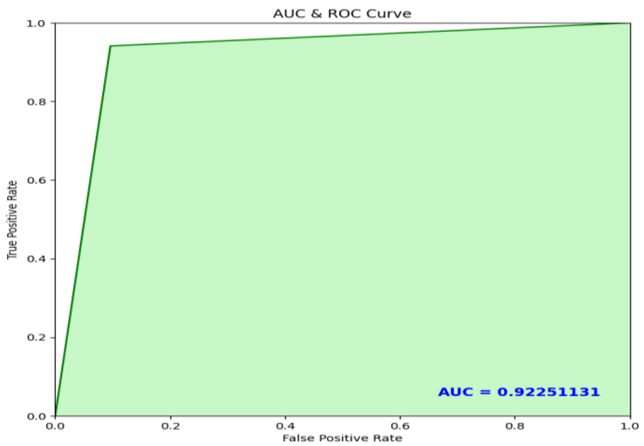


Figure 6: Performance of RF model using the ROC curve

classification thresholds. AUC value 1 specifies perfect classification and 0.5 relates to imperfect models (Walter, 2005; Nguyen et al., 2019). The model accuracies with AUC values are interpreted as; 0.6 to 0.7 poor, 0.7 to 0.8 good, 0.8 to 0.9 very good and 0.9 to 1 excellent (Pourghasemi et al., 2017). Performance of the RF model based on the validation datasets (30%) of the study area is shown in figure 6. Based on AUC matric (0.922) performance of RF Model is excellent for recognizing pattern of flood susceptibility. Results of other statistical measures (Accuracy, Precision, Recall and F Score) used to validate flash flood modeling are summarized in Table 4.

Name	Equation	Optimum value	Meaning	Performance
Accuracy	TP+TN / TP+FP+TN+FN	1	Proportional measure of the number of predictions over all predictions	0.925
Precision	TP / TP+FP	1	Proportion of correct positive predictions	0.903
Recall	TP / TP+FN	1	Proportion of actual positives identified correctly	0.921
F score	2*(Precision*Recall) / Precision + Recall	1	Harmonic mean of Precision and Recall	0.911

Table 4: Performance of RF model using statistical measures on the validation dataset

Based on modeling performance results it can be stated that RF model could be effective tool for flash flood susceptibility assessment in the region.

4. CONCLUSIONS

Flash floods are one of the most common forms of natural disaster in the Himalayan region that cause extensive damage to life and property. Therefore, effective approaches are needed to delineate the most sensitive locations of the region in order to minimize losses to this disaster. Machine learning is one of the approaches that is increasingly being used for hazard susceptibility predictions. The present study aims to investigate effectiveness of Random Forest algorithm based on real-world dataset of Indian Central Himalayans Region that is one of the most Flash flood prone areas in the Himalayan region. A flash flood inventory with 400 locations (200 flash flood locations and 200 unflooded locations) and eight conditioning factors namely elevation, slope, aspect, profile curvature, distance from river, rainfall, lithology and LULC were used for Random Forest Modeling. The validation results indicate an AUC of 0.922 and other performance measures including accuracy, precision, Recall and F Score also confirmed results of the AUC value. The study concludes that RF model approach has great potential for flood susceptibility assessment in the region. There is further space for improvement of model performance and model can also be applied in the other Himalayan regions by considering greater number of flash flood locations and optimum conditioning factors based on topographical, geological, hydrological and meteorological conditions. RF Modeling can significantly contribute to improving the understanding of planners to review their conservation plan for future floods in the Himalayan region.

Authors' contributions: All authors contributed equally

Conflict of interest: The authors declare no conflict of interest

REFERENCES

[1] Abedi, R., Costache, R., Shafizadeh-Moghadam, H., & Pham, Q. B. (2022). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, *37*(19), 5479-5496.

[2] Abedini, M., Ghasemian, B., Shirzadi, A., Shahabi, H., & Chapi K. (2019). A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto International*, *34* (13), 1427–1457.

[3] Amit, Y. & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural computation*, *9* (7), 1545–1588

[4] Aryal, S.K., Mein, R.G.& O'Loughlin, E.M. (2003). The concept of effective length in hill slopes: assessing the influence of climate and topography on the contributing areas of catchments. *Hydrological Processes*, *17*(1), 131–151

[5] Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Melesse, A.M. & Mosavi, A. (2020). Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sensing*, *12*(21), 3568.

[6] Bhambri, R., Mehta, M., Dobhal, D.P., Gupta A.K., Pratap, B., Kesarwani, K. & Verma, A. (2016). Devastation of Kedarnath (Mandakini) valley, Garhwal Himalaya, during 16-17 June 2013: A remote sensing and ground-based assessment, *Natural hazards*, *80* (3), 1801-1822.

[7] Bordbar, M., Aghamohammadi, H., Pourghasemi, H.R. & Azizi, Z. (2022). Multi-hazard spatial modeling via ensembles of machine learning and meta-heuristic techniques. *Scientific Reports*, *12*(1), 1451.

[8] Breiman, L. (1996a). Bagging predictors. *Machine Learning*, *24*,123–140.

[9] Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

[10] Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang. X. & Ahmad, B.B. (2019). Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology*, *575*, 864–873.

[11] Dimri, A. P., Chevuturi, A., Niyogi, D., Thayyen, R. J., Ray, K., Tripathi, S. N. & Mohanty, U. C. (2017). Cloudbursts in Indian Himalayas: a review. *Earth-Science Reviews*, *168*, 1-23.

[12] Fawagreh, K., Gaber, M. M. & Elyan, E. (2014). Random forests: from early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal*, *2*(1), 602-609.

[13] Goetz, J. N., Brenning, A., Petschko, H. & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & geosciences*, *81*, 1-11.

[14] Ho, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on *document analysis and recognition*, Montreal, Quebec, Canada 1: 278-282. New York City, NY: IEEE.

[15] Joshi, V. & Kumar, K. (2006). Extreme rainfall events and associated natural hazards in Alaknanda valley, Indian Himalayan region. *Journal of Mountain Science*, *3*, 228-236.

[16] Khanduri, S. (2020). Cloudbursts over Indian sub-continent of Uttarakhand Himalaya: A traditional habitation input from Bansoli, District-Chamoli, India. *International Journal of Earth Sciences Knowledge and Applications*, *2*(2), 48-63.

[17] Khanduri, S. (2021). Formation and failure of natural dams in Uttarakhand Himalaya: An Observation from Lwarkha, Chamba Tahsil of Tehri Garhwal District, India. *International Journal of Earth Sciences Knowledge and Applications*, *3*(1), 12-22.

[18] Khanduri, S. (2022). Rain-Induced Slope Instability: Case Study of Monsoon 2020 affected Villages in Pithoragarh District of Uttarakhand, India. *International Journal of Earth Sciences Knowledge and Applications*, *4* (1), 1-18.

[19] Kumar, A., Gupta, A. K., Bhambri, R., Verma, A., Tiwari, S. K. & Asthana, A. K. L. (2018). Assessment and review of hydrometeorological aspects for cloudburst and flash flood events in the third pole region (Indian Himalaya). *Polar Science*, *18*, 5-20.

[20] Modrick, T.M. & Georgakakos, K.P. 2015. The character and causes of flash flood occurrence changes in mountainous small basins of Southern California under projected climatic change. Journal of Hydrology: *Regional Studies*, *3*, 312–336.

[21] Nguyen, P. T., Tuyen, T. T., Shirzadi, A., Pham, B. T., Shahabi, H., Omidvar, E. & Bui, D. T. (2019). Development of a novel hybrid intelligence approach for landslide spatial prediction. *Applied Sciences*, *9*(14), 2824.

[22] Nguyen, V. V., Pham, B. T., Vu, B. T., Prakash, I., Jha, S., Shahabi, H. & Tien Bui, D. (2019). Hybrid machine learning approaches for landslide susceptibility modeling. *Forests*, *10*(2), 157.

[23] NIDM (2015). Uttarakhand Disaster 2013. National institute of disaster management (Ministry of home affairs, Govt of India, New Delhi.

[24] Maestre, F. T. & Cortina, J. (2002). Spatial patterns of surface soil properties and vegetation in a Mediterranean semi-arid steppe. *Plant and soil*, *241*, 279-291.

[25] Parkash, S. (2015). A study on flash floods and landslides disaster on 3rd August 2012 along Bhagirathi Valley in Uttarkashi District, Uttarakhand. World Centre of Excellence on Landslide Disaster Reduction, National Institute of Disaster Management, Ministry of Home Affairs, Government of India.

[26] Pham, B. T., Jaafari, A., Prakash, I., Singh, S. K., Quoc, N. K. & Bui, D. T. (2019). Hybrid computational intelligence models for groundwater potential mapping. *Catena*, *182*, 104101.

[27] Pham, B. T., Phong, T. V., Nguyen, H. D., Qi, C., Al-Ansari, N., Amini, A. & Tien Bui, D. (2020). A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes, and logistic model tree for flash flood susceptibility mapping. *Water, 12*(1), 239.

[28] Pourghasemi, H. R., Yousefi, S., Kornejady, A. & Cerdà, A. (2017). Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Science of the Total Environment*, *609*, 764-775.

[29] Quinn, P. E., Hutchinson, D. J., Diederichs, M. S. & Rowe, R. K. 2011. Characteristics of large landslides in sensitive clay in relation to susceptibility, hazard, and risk. *Canadian Geotechnical Journal*, *48*(8), 1212-1232.

[30] Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A. & Shufeng, T. (2019). Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. *Earth Systems and Environment*, *3*, 585-601.

[31] Rahmati, O., Pourghasemi, H. R.& Zeinivand, H. 2016. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International*, *31*(1), 42-70.

[32] Rana, M. S.& Mahanta, C. 2023. Spatial prediction of flash flood susceptible areas using novel ensemble of bivariate statistics and machine learning techniques for ungauged region. *Natural Hazards*, *115*(1), 947-969.

[33] Rautela, P. (2016). 16/17 June 2013 disaster of Uttarakhand, India and lessons learnt. *Disaster- Response and Management Journal*, *4*, 42-70.

[34] Rusk, J., Maharjan, A., Tiwari, P., Chen, T. H. K., Shneiderman, S., Turin, M.& Seto, K. C. (2022). Multi-hazard susceptibility and exposure assessment of the Hindu Kush Himalaya. *Science of the total environment*, *804*, 150039.

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

[36] SANDRP - South Asia Network on Dams, Rivers and People. https://sandrp.in

[37] Sati, S. P., Sharma, S., Sundriyal, Y. P., Rawat, D. & Riyal, M. (2020). Geo-environmental consequences of obstructing the Bhagirathi River, Uttarakhand Himalaya, India. *Geomatics*, *Natural Hazards and Risk*, 11(1), 887-905.

[38] Sati, V. P. (2007). Environmental impacts of debris flows - A case study of the two debris-flow zones in the Garhwal Himalaya. Debris - Flow Hazards Mitigation: *Mechanics, Prediction, and Assessment,* Chen & Major, eds, 715: 723.

[39] Sati, V. P. and Kumar, S. (2022). Environmental and economic impact of cloudburst-triggered debris flows and flash floods in Uttarakhand Himalaya: a case study. *Geoenvironmental Disasters*, *9* (1), 5.

[40] Singh, G. & Pandey, A. (2022). Hybrid ensemble modeling for flash flood potential assessment and susceptibility analysis of a Himalayan River catchment. *Geocarto International*, *37*(25), 9132-9159.

[41] Singh, H., Varade, D. & Mishra, P. K. (2022). Cloudburst events in the Indian Himalayas: a historical geospatial perspective. In *International Handbook of Disaster Research* (pp.1-21). Singapore: Springer Nature Singapore.

[42] Vishwanath, V.H. & Tomaszewski, B.M. 2018. Flood Hazard, Vulnerability, and Risk Assessment for Uttarakhand state in India. *Proceedings of the 15th ISCRAM Conference*, Rochester, NY, USA

[43] Vojtek, M. & Vojteková, J. (2019). Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water*, *11*(2), 364.

[44] Walter, S. D. (2005). The partial area under the summary ROC curve. *Statistics in medicine*, *24* (13), 2025-2040.

[45] Wang, X., Kinsland, G., Poudel, D. & Fenech, A. 2019. Urban flood prediction under heavy precipitation. *Journal of Hydrology*, *577*, 123984.

[46] Youssef, A. M., Pradhan, B. & Sefry, S. A. (2016). Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. Environmental *Earth Sciences*, *75*(1), 12.