



A novel hybrid of meta-optimization approach for flash flood-susceptibility assessment in a monsoon-dominated watershed, Eastern India

Dipankar Ruidas¹ · Rabin Chakraborty¹ · Abu Reza Md. Towfiqul Islam² · Asish Saha¹ · Subodh Chandra Pal¹

Received: 27 September 2021 / Accepted: 29 January 2022 / Published online: 21 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The exponential growth in the number of flash flood events is a global threat, and detecting a flood-prone area has also become a top priority. The flash flood-susceptibility mapping can help to mitigate the worst effects of this type of risk phenomenon. However, there is an urgent need to construct precise models for predicting flash flood-susceptibility mapping, which can be useful in developing more effective flood management strategies. In this present research, support vector regression (SVR) was coupled with two meta-heuristic algorithms such as particle swarm optimization (PSO) and grasshopper optimization algorithm (GOA), to construct new GIS-based ensemble models (SVR–PSO and SVR–GOA) for flash flood-susceptibility mapping (FFSM) in the Gandheswari River basin, West Bengal, India. In this regard, 16 topographical and environmental flood causative factors have been identified to run the models using the multicollinearity (MC) test. The entire dataset was divided into 70:30 for training and validating purposes. Statistical measures including specificity, sensitivity, PPV, NPV, AUC–ROC, kappa and Taylor diagram have been employed to validate adopted models. The SVR-based factor importance analysis was employed to choose and prioritize significant factors for the spatial analysis. Among the three modeling approaches used here, the ensemble method of SVR–GOA is the most optimal (specificity 0.97 and 0.87, sensitivity 0.99 and 0.91, PPV 0.97 and 0.86, NPV 0.99 and 0.91, AUC 0.951 and 0.938 in training and validation, respectively), followed by the SVR–PSO (specificity 0.84 and 0.84, sensitivity 0.87 and 0.86, PPV 0.85 and 0.82, NPV 0.87 and 0.87, AUC 0.951 and 0.938 in training and validation, respectively) and SVR (specificity 0.80 and 0.77, sensitivity 0.93 and 0.89, PPV 0.82 and 0.77, NPV 0.91 and 0.89, AUC 0.951 and 0.938 in training and validation, respectively) model. The result shown that 40.10 km² (10.99%) and 25.94 km² (7.11%) areas are under very high and high flood-prone regions, respectively. This produced reliable results that can help policymakers at the local and national levels to implement a concrete strategy with an early warning system to reduce the occurrence of floods in a region.

Keywords Artificial intelligence · Flash flood · Multicollinearity · Optimization · Specificity

Introduction

Every year, various natural disasters such as earthquakes, landslides, and floods cause massive loss of life and property worldwide (Band et al. 2020a). Flood is one of the most dangerous natural hazards, responsible for thousands of deaths, ecosystem damage, communication system disruption, and billions of dollars in economic losses each year (Luu et al. 2018; Aerts et al. 2018; Khosravi et al. 2018; Pham et al. 2021a). Many regions across the world are affected by water excess-related natural events through snow melting or intensified rainfall (Burgan and Icaga 2019), which can be dangerous if they occur unexpectedly. As a natural phenomenon, flash floods occur most commonly in river basins smaller

This article is part of a Topical Collection in Environmental Earth Sciences on Recent Advances in Environmental Sustainability, guest edited by Peiyue Li.

✉ Subodh Chandra Pal
geo.subodh@gmail.com

¹ Department of Geography, The University of Burdwan, Bardhaman 713104, West Bengal, India

² Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh

than 200 km² (Costache and Tien Bui 2020). In recent times, climate change and LULC change have been identified as the primary causes of flash floods (Sekac et al. 2015; Kanani-Sadat et al. 2019). It could happen due to heavy rainfall, sudden ice melting on the upper ridge, or poor infrastructure in the flood risk zone. Several hydrological, meteorological, geomorphological, and flood protection failures are to blame for this hydrological event, as are various human activities such as encroachment on the river catchment, deforestation, and dam development, which increase the frequency of floods (Roy et al. 2020). Climate change has also exacerbated the destructive impact of flash floods (Bubeck and Thielen 2018). This hydrological phenomenon causes catastrophic mudslide and debris flow that is characterized by short response time and results in massive socioeconomic loss all over the world (Peduzzi 2017). The flood assessment studies of Wasko and Sharma (2017) showed that flash flood events of 2013 caused heavy damages with an estimation of more than 50 billion US dollars globally. Additionally, they also forecast it will double within the next 20 years. In the case of flood events, it is estimated that by 2050 the expected potential loss will reach 1 trillion US dollars due to the rapid increase of anthropogenic activities and extreme weather events (Hartnett and Nash 2017). According to the World Meteorological Organization (2016), flash flood has a high mortality rate, accounting for over 5000 lives lost each year, and 85% of flooding cases that result in around 1 million deaths worldwide were due to the deadliest flood events during the 1972–2012 time period. The reports of WHO display that approximately 2 billion people are affected globally between 1998 and 2017. Thus, flash flood event has got substantial attention particularly in monsoon-dominated region. Therefore, flood researchers around the world are constantly looking for new ways to mitigate the hazardous effects of flooding in high-risk areas. Hence, the management of floods is yet subject to a thorough study.

In the modern era, no country is immune to flood hazards, and Asian countries are especially vulnerable due to their dense human populations along river floodplains (Hens et al. 2018; Pham et al. 2021b). Just in the twenty-first century, India has experienced several devastating flash flood events such as the Mumbai flood (2005), Assam flood (2012), Uttarakhand flood (2013), Jammu Kashmir flood (2014), Chennai flood (2015), Kerala flood (2018), and Uttarakhand flood at Chamoli district (2021). All these flood events have caused massive loss of lives and property in India. One of the major causes of increasing flood events and losses is a lack of adequate studies and information about flood-prone areas and lack of flood prevention ideas, and it is evident that flood events cannot be prevented but they can be predicted using appropriate methodologies and analysis (Khosravi et al. 2016). The paradigm has recently shifted from “flood protection” to “flood risk management” (Scheuer et al. 2011;

Varu et al. 2020). Hence, to manage the flood risk, flash flood-susceptible zonation is a crucial tool that plays a significant role in flood risk reduction through its predictive flood-prone area identification (Tehrany et al. 2015; Youssef et al. 2015; Pal et al. 2018; Band et al. 2020b; Chowdhuri et al. 2020; Malik and Pal 2021). Thus, West Bengal also experienced significant numbers of severe flood events (1978, 1991, 2000, 2006, 2013, 2017) and many small-scale regional flood occurrences that caused a notable number of deaths and property loss. According to the disaster management department of West Bengal, Bankura district is under a relatively flood-prone district due to two main rivers, namely, Dwarakeswar and Gandheswari. Among them, the Gandheswari River flood has a great impact on Bankura district due to its approximately central location of Bankura town, which has experienced several flood events in 2016, 2018, and 2021. Therefore, in this study, we attempt to make a flash flood-susceptible map (FFSM) to identify specific flood-prone regions of the Gandheswari River basin, which can be very helpful for proper strategy to reduce the severity of flood occurrence. Flash flood events may occur in any region as a result of a variety of factors such as geological, hydrological, and meteorological factors. The analysis of quantitative characteristics of all those factors in the considered basin area is important to know the hydrological behaviour of this river. Using the combined RS–GIS technique, 16 flood conditioning parameters were identified including, geology, geomorphology, soil texture, slope, profile curvature, plan curvature, elevation, aspect, distance to the river, rainfall, land use, and land cover (LULC), and other hydrological parameters to make FFSM in the considered basin area.

In proper flood disaster management, it is crucial for the identification of real-time flood inundation maps (Chang et al. 2018) and to forecast this systematically and scientifically. Numerous approaches have been used that help to forecast the different features of flood events: the statistical correlation of flood characteristics with the physical properties of the watershed (Wheater et al. 1993) and various evidential methods such as the multi-criteria decision analysis (MCDA) and analytical hierarchy process (AHP) used to identify the flood-susceptibility zones. But to handle the complexities of data structure and to make high accuracy in recent time, the machine learning (ML) approach has gained more attention from researchers across the globe (Tehrany et al. 2015; Khosravi et al. 2018; Band et al. 2020b). Several ML algorithms such as random forest (RF), logistic regression (LR), support vector machine (SVM), artificial neural network (ANN) and various ensemble approaches have been used to identify flood-susceptibility mapping (Luu et al. 2018; Bui et al. 2018; Costache and Tien Bui 2020; Malik et al. 2020; Saha et al. 2021a). Hybrid and combined modeling approaches are beneficial for mapping

flash flood susceptibility. Most of the researchers have used multiple ML approaches rather than a single ML approach to obtain better results (Band et al. 2020b). A literature review revealed that the application of pragmatic kernel-based models, such as support vector regression (SVR), is one of the most popular tools for mapping flash flood susceptibility, especially in data-limited regions such as West Bengal, India (Paul et al. 2019; Roy et al. 2020). However, the support vectors identified the optimal hyperplane; a certain kernel function was adopted to convert the input dataset into two distinct categories and also avoid developing a linear association between two training and validation datasets (Choubin et al. 2019). Recently, more advanced algorithm of short-term hybrid memory (LSTM) and ant lion optimizer (ALO) modeling has been used in streamflow analysis with good accuracy (Yuan et al. 2018). Dynamic evolving neural fuzzy inference system (DENFIS) model was used in evaluating the performance of relationship between the capturing rainfall-runoff and rainfall-water level (Chang et al. 2019). Furthermore, group method of data handling (GMDH) approach has been applied in flood (Dodangeh et al. 2020) and landslide (Panahi et al. 2022) susceptibility analysis. Studies also show that least squares support vector machine (LSSVM) and gravitational search algorithm (GSA) have been widely used in various prediction analysis. Among them Yuan et al. (2015) used LSSVM and GSA model in short-term wind prediction analysis and their outcomes gives good prediction result. The hybrid algorithm of extreme learning machine (ELM) and PSOGWO was used in monthly runoff-based streamflow prediction analysis in Mangla watershed of Pakistan (Adnan et al. 2021). All of the above-mentioned methods have been used in several research studies and their outcomes are also very good. In addition to this, state-of-the-art meta-heuristic algorithm has been widely used by several researchers across the world for prediction analysis in different fields. In this study, flash flood-susceptibility mapping was performed using SVR, coupled with the particle swarm optimization (PSO) method (SVM-PSO) and grasshopper optimization algorithm (GOA) method (SVM-GOA) as two meta-heuristic algorithms, to optimize the SVR model. The above selected learning algorithms have great capacity in the field of prediction analysis and is more suitable for large dataset analysis; therefore, we have selected these models. State of the art generally refers to the highest level of achievement in a specific purpose, i.e. performance analysis. State-of-the-art algorithm is used in the field of landslide susceptibility (Mehrabi et al. 2020), groundwater potential zoning (Arabameri et al. 2021), flood studies (Arora et al. 2021) and many more. Overall, the predictive accuracy of flash flood susceptibility has been enhanced substantially by applying these ensemble models. However, no common consensus has still been established on choosing the optimal model for flash flood-susceptibility mapping; therefore,

novel models are required and should be verified. Nevertheless, our study intended to answer the two research questions: firstly, what are the influential variables attributing to flash flood susceptibility? Secondly, what are the areas flooded most frequently? To answer the two questions, the primary goal was to (i) identify the influential variable for flash flood events using the multicollinearity test in the Gandheswari River basin and (ii) to prepare FFSM using state-of-the-art ensemble and standalone machine learning methods (SVR, SVR-PSO, and SVR-GOA) and (iii) compare the results to produce the region's accurate FFSM. The novel aspect of our study is that it couples with the ensemble (SVR-PSO and SVR-GOA) for FFSM. As we know, the ensemble approach is the best for any kind of susceptibility analysis than any single model. Therefore, an ensemble of SVR-GOA and SVR-PSO was chosen in this study to predict flood-susceptibility analysis. As our study revealed that the used ensemble of SVR-PSO and SVR-GOA models for predicting flood-susceptibility mapping is the first, and those will be significant modeling approaches in identifying flash flood-susceptible zone. The reason for choosing the SVR is that its computational complexity does not depend on the dimensionality of the input space, along with its high prediction accuracy with complex capacity. Although other learning models also have good analysis capacity, SVR has been used in flood-susceptibility analysis and gives optimal prediction, so here we have chosen this model. Additionally, PSO and GOA have been widely used in several fields of prediction analysis due to their unique performance analysis, so here we have selected this model for ensemble with SVR. Lastly, it is worth mentioning that the proposed model is limited to not only flash flood susceptibility, but also can be used in any natural hazards.

Material and methods

Study area

The designated area of the study to implement the proposed method is the Gandheswari River basin which is located in the Bankura district of West Bengal, India, with an absolute location of 23°13'28"N to 23°30'25"N latitude and 86°53'13"E to 87°07'30"E longitude (Fig. 1). This river basin covers approximately 364.9km², is elongated in shape and structurally controlled (Mondal and Mistri 2015). Gandheswari is a tributary of the Dwarakeswar River, which flows through the Bankura district. From the north-western to the south-eastern part of the district, the Gandheswari River flows through four blocks: Bankura, Saltora, Chhatna, and Gangajalghati, before joining the Dwarakeswar River. Geologically, this area is dominated by Chotonagpur gneissic complex, and pediment pediplain complex is the significant

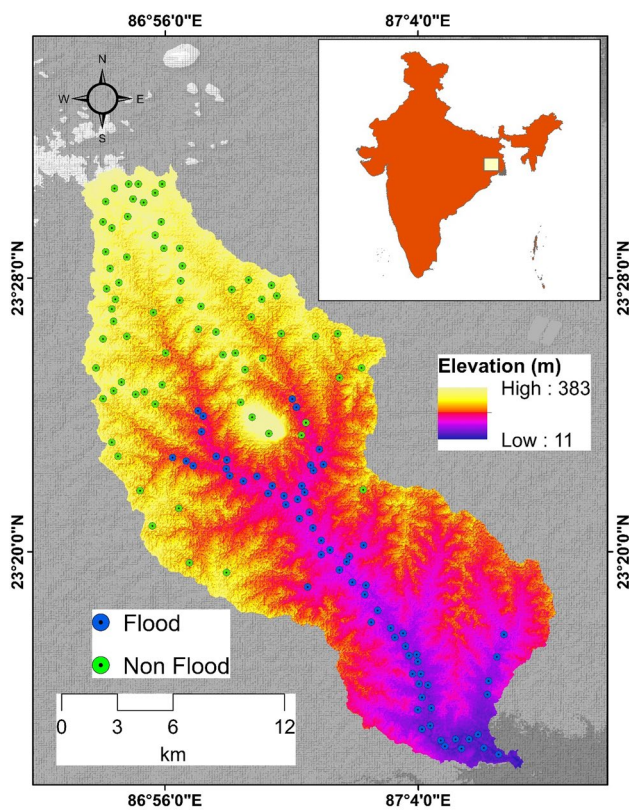


Fig. 1 Location map of the study area and flood inventory map

geomorphic characteristic with an 11–383 m elevation range (GSI). The average rainfall in the study area is 1400 mm, and it is located in a monsoonal climatic region. Because of the SW monsoon, a considerable amount of precipitation falls during July and August. In the last few years, substantial changes in LULC along the river channel make this region vulnerable to flash floods; population and settlement density in this region have increased day by day due to socioeconomic development in this region that significantly increased encroachment toward the river channel that increased pressure on river flow. These geographical conditions and anthropogenic pressure are responsible for accelerating flash flood events. The disaster management department of West Bengal identified the Bankura district as a relatively flood-prone district; as a result, the identification of flash flood-prone region is more essential and can help to take necessary steps to reduce the adverse effect of flash flood in this river basin.

Methodology

The exact FFSM and its level of accuracy are determined by the size and availability of the data, as well as the methods used to create this map. It is necessary to develop a proper methodological framework (Fig. 2) that aids in determining

flood-prone areas. In this study, we adopted the following steps for this assessment:

- (I) A flood inventory map was prepared based on 140 flood points; of those, 70 were flood areas and 70 non-flood areas.
- (II) Sixteen appropriate flash flood causative factors were identified based on an extensive literature survey and local geo-environmental conditions.
- (III) Multicollinearity test was adopted among various flash flood causative factors by using variance inflation factor (VIF) and tolerance (TLT) methods.
- (IV) Support vector regression (SVR), particle swarm optimization (PSO), and grasshopper optimization algorithm (GOA) models were used to model the FFSM.
- (V) Support vector regression (SVR), ensemble method of SVR–PSO, and SVR–GOA were used in flash flood-susceptibility mapping.
- (VI) The projected result of these models was validated by six statistical techniques such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and receiver operating characteristics (ROC) – AUC, kappa coefficient analysis and one graphical technique of Taylor diagram.

Flood inventory map

Flash flood inventory mapping plays a significant role in establishing the relationship between flood and various causative factors. It is the most important requirement in FFSM (Arabameri et al. 2019). Various methods are used to make it, but it depends on several parameters such as the environmental condition of the study area, research purpose, and access of RS–GIS data (Pradhan 2013; Arabameri et al. 2019). To create an FFSM, first, we gather the necessary data and create a spatial database. It is necessary to provide the necessary data with high accuracy. The inventory area has previously been affected by torrential flash floods and can detect areas where flash floods may occur in the future. This area is also prone to rapid surface runoff, which contributes to flash flood events on the downslope. It is justified by the fact that the accuracy of FFSM mostly depends on proper inventory mapping, which is based on extensive historical data analysis (Band et al. 2020b; Saha et al. 2021a). In this study, a total of 140 points were identified, among them, 70 points depict flood areas and 70 were non-flood areas. Google Earth's image of the time of flood events and the extensive field inventories were the main basis of flood area selection. In the same way, non-flood areas have been randomly selected using ArcGIS software.

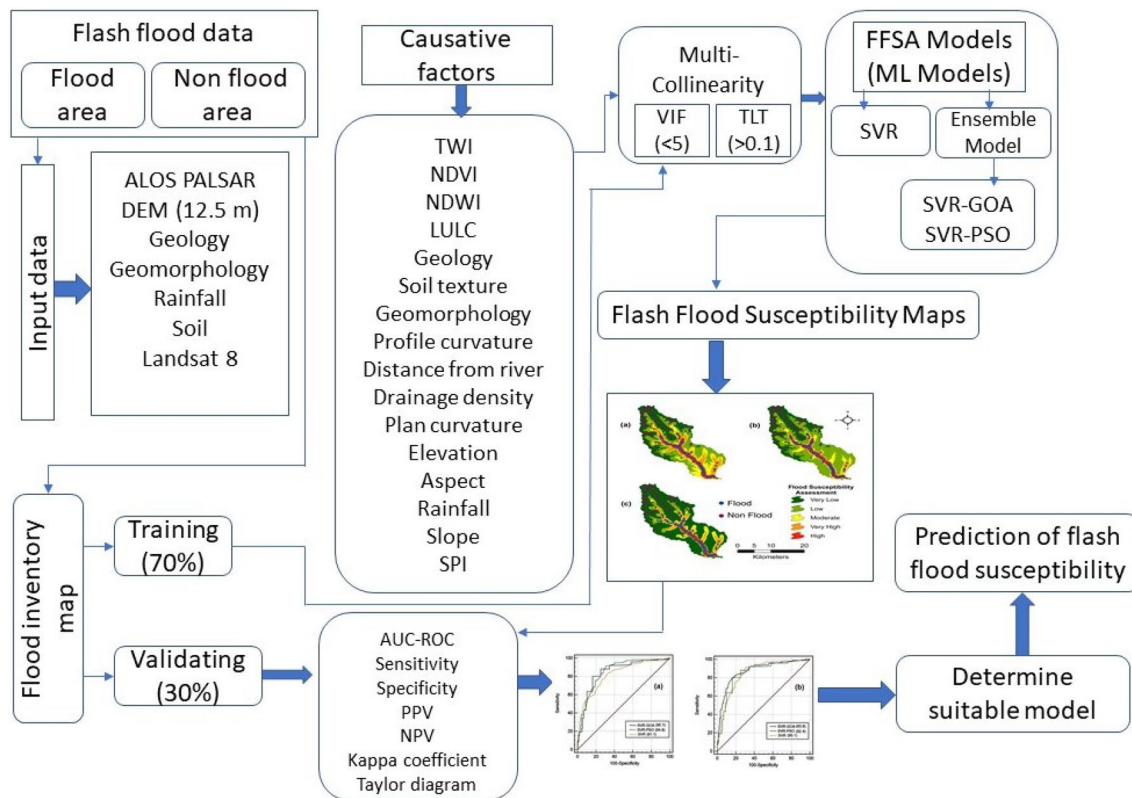


Fig. 2 Methodological framework of FFSA

Flash flood causative factors

As a complex natural incident, the flash flood in a particular watershed does not only depend on the hydro-meteorological factors, but also on the geomorphological condition of that watershed (Tien Bui et al. 2019); Thus, it is necessary to identify the flash flood causative factors and their relationship with flash flood events (Khosravi et al. 2016). It means the sensitivity analysis and the relation of each causative factor should be estimated (Twele et al. 2016; Arabameri et al. 2019). In the present study, based on the environmental condition of the study area, historical flood events, and various studies (Khosravi et al. 2016, 2018; Chapi et al. 2017; Arabameri et al. 2019; Band et al. 2020b; Chowdhuri et al. 2020; Roy et al. 2020; Malik and Pal 2021), 16 flood causative factors were used. Furthermore, the 16 flood-susceptibility factors were chosen for this study area depending upon the various local topo-hydrological, climatological, geological, environmental and geomorphological condition. In flood-susceptibility analysis, local conditioning factors are very important for accurate analysis and to understand the flood condition. Alongside, various literature reviews have also been conducted for the selection of appropriate flood conditioning factors. These factors were geology, geomorphology, soil texture, normalized difference vegetation

index (NDVI), normalized difference water index (NDWI), land use and land cover (LULC), profile curvature, rainfall, distance from the river, plan curvature, stream power index (SPI), drainage density, aspect, topographic wetness index (TWI), elevation and slope, which have direct and indirect relation with flash flood events in the considered area.

All these considered parameters were prepared from different sources and by using the GIS environment including geology and geomorphology maps prepared from the Geological Survey of India (GSI) and soil texture map collected from the National Bureau of Soil Survey and Land Use Planning (NBSS and LUP), Government of India. The rainfall map of this study area was prepared by the inverse distance weighting (IDW) method with the help of the Indian Meteorological Department rainfall data. ALOS PALSAR DEM (<https://asf.alaska.edu/>) with 12.5-m resolution helps to extract slope, elevation, aspect, plan curvature, and profile curvature in Arc GIS 10.4 where SPI and TWI are developed by SAGA-GIS. ALOS PALSAR also helps to extract distance from the river and drainage density in Arc hydro environment. LULC, NDVI, and NDWI map were prepared from Landsat 8 (<https://earthexplorer.usgs.gov/>).

Elevation was a significant causative factor of a flash flood that controls the natural flow of a river, and it can regulate the microclimatic conditions of a region that leads

to development in plant coverage which has a significant impact on runoff formation (Tien Bui et al. 2020). Higher elevation was usually safe from flash floods, whereas lower elevation was highly vulnerable to flash flood events (Malik et al. 2020; Islam et al. 2021; Malik and Pal 2021), so it has an inverse association with height. The elevation ranges from 11 to 383 m (Fig. 3a) in the present study and the northern and southern portions are characterized with higher and lower elevations, respectively.

Slope defines the intensity and magnitude of water percolation and water accumulation (Choubin et al. 2019). According to Florinsky (2016), the slope of the Earth's surface can influence several hydrogeological processes such as infiltration and runoff that have a great impact on the occurrence of the flash flood. An adverse relationship is present between slope and flood. The heavy rainfall in the steep slope region helps to originate high energy in the surface runoff that leads to faster flow toward low land, resulting in a flash flood (Ngo et al. 2021). Very low slope with high surface rainfall causes low surface runoff which leads to flooding (Rahmati et al. 2016). On the other hand, a high slope is prone to high runoff that is less susceptible to flood. The slope ranges from 0° to 43.18° (Fig. 3b); the higher slope is found in the central region, and the lower part along the river of the study area is characterized with a lower slope.

Aspect of slope is an angle between the north and horizontal direction and has substantial influence on soil water (Ragab et al. 2003). Thus, the aspect is considered an important flood-susceptibility forecasting parameter (Zhao et al. 2018). This shows the orientation of the slope (Choubin et al. 2019). In this study, the slope aspect varies from -1 to 359.23 (Fig. 3c).

TWI mainly indicates the moisture condition of the soil, saturation level, and water depth of topography (Band et al. 2020b), influence to plant diversities and distribution. This can closely regulate the depletion and spatial spreading of surface runoff (Bui et al. 2018). Generally, high TWI values indicate more vulnerability to flooding compared to low TWI values. In the current study, the range is from 6.94 to 27.78. (Fig. 3d); high TWI is found mainly along the river that implies this region is significantly prone to flash floods. TWI is calculated from the following equation:

$$TWI = \ln \left(\frac{A_s}{\tan(\text{slope})} \right), \quad (1)$$

where A_s represents the area of the basin (km^2) and $\tan(\text{slope})$ is the slope of the basin.

Stream power index (SPI) can be described as the erosive power and surface runoff of an area (Malik et al. 2020). High SPI represents the high surface runoff, whereas the low SPI represents the low surface runoff of a river basin. Low SPI with high intensity of rainfall makes the area flood

susceptible. In this study, SPI ranges from 0.34 to 19.75 (Fig. 3e); the higher value location of SPI implies that the immediate location of the stream is highly vulnerable to flash floods. The following equation has been used to calculate SPI:

$$SPI = A_s * \tan(\text{slope}), \quad (2)$$

where A_s describes the catchment area (km^2) and $\tan(\text{slope})$ is the slope of the catchment.

Drainage density plays a significant role in FFSM which is determined by the drainage length per unit area (Horton 1932); this has a crucial role in storage and transfer of water of a watershed, and it is evident that higher river density causes higher flood events (Khosravi et al. 2018). A drainage system is the natural flow of the runoff, and when the channelized flow exceeds its channel potential, it makes the area flooded. Figure 3f shows the drainage condition of the study area where the middle to a lower portion of the study area is characterized by higher drainage density.

Distance to the river can easily describe which region is more susceptible to flood and which is less (Malik et al. 2020). The high velocity of stream flow occurs at the time of heavy and stormy runoff period and at that time it reaches its maximum limit and makes the area flooded. The area which is close to the river is most vulnerable to flood compared to the high distant area from the river (Fig. 3g).

Plan curvature is the terrain's curvature concerning the slope direction which plays a significant role in predicting flood susceptibility. A positive value represents a convex profile where the flat profile is described by zero and a negative value indicates a concave profile; the plan curvature of the study area ranges between -3.18 and 2.96 . (Fig. 3h). Plan curvature affects the runoff pattern; therefore, it has valuable insight into FFSM.

Profile curvature is the surface curvature against the cliff which is very close to the plan curvature. It has a substantial influence on surface runoff and is a significant indicator of flood events (Arabameri et al. 2019). The profile curvature of this area is between -3.32 and 2.61 (Fig. 3i).

Rainfall is a significant parameter in the magnitude of the flash flood events, which has a direct relation between the intensity of rainfall and flash flood. According to Guardiola-Albert et al. (2020) and Gaume et al. (2004), high-intensity rainfall in small watershed regions triggers flash flood occurrence. The rainfall amount varies from 515 to 537 mm (Fig. 3j) and a higher amount of rainfall is mainly found in the southern portion of the study area which mainly occurs due to SW monsoon. As a primary driving force, the short duration, high intensity, high frequency, and small spatial distribution of rainfall make the watershed more vulnerable to flooding with the association of other flood-causing factors.

Fig. 3 Flood causative factors: **a** elevation, **b** slope, **c** aspect, **d** topographic wetness index TWI, **e** stream power index SPI, **f** drainage density, **g** distance to river, **h** plan curvature, **i** profile curvature, **j** rainfall, **k** NDVI, **l** NDWI, **m** geology, **n** geomorphology, **o** soil texture and **p** LULC

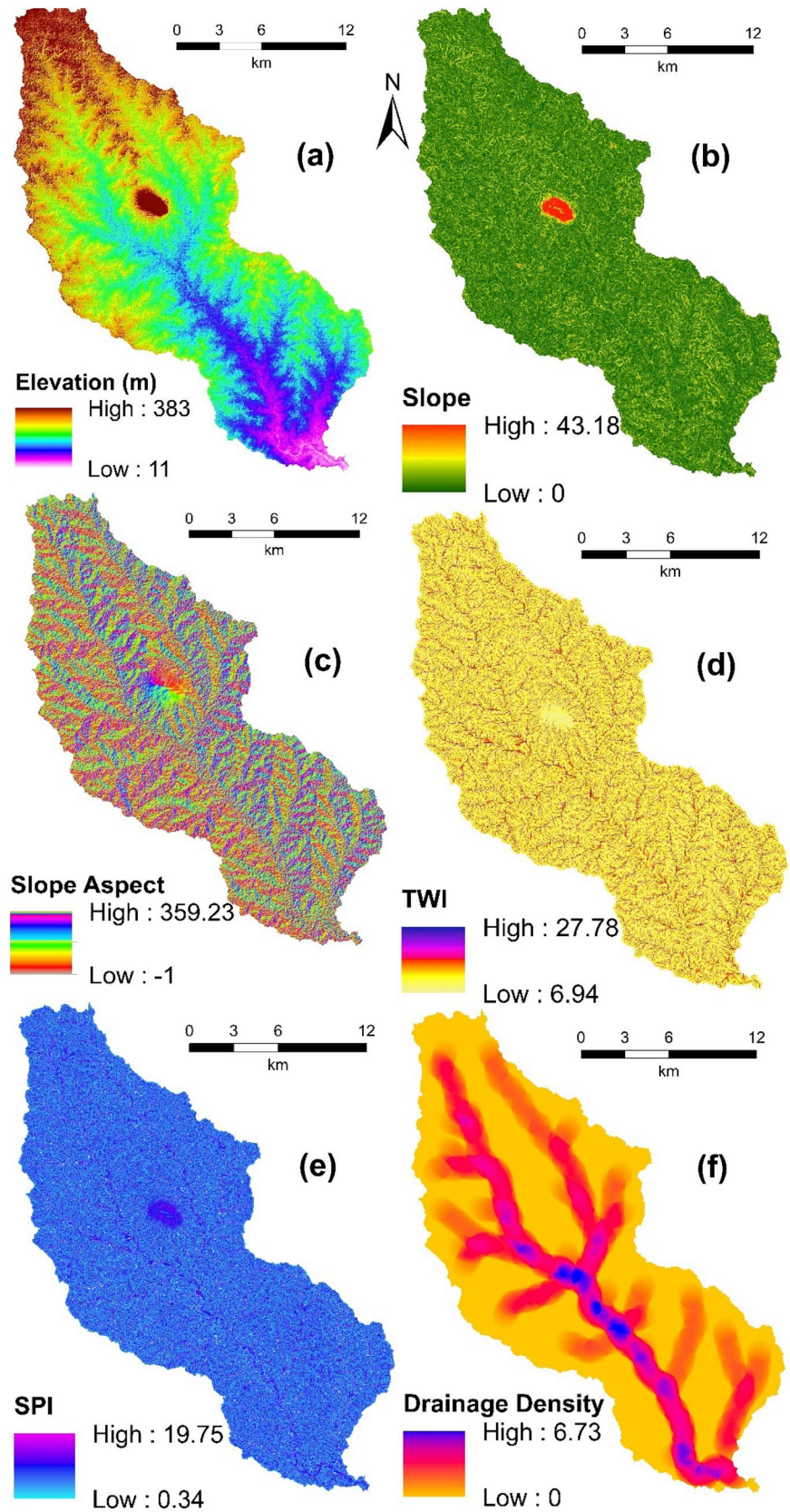


Fig. 3 (continued)

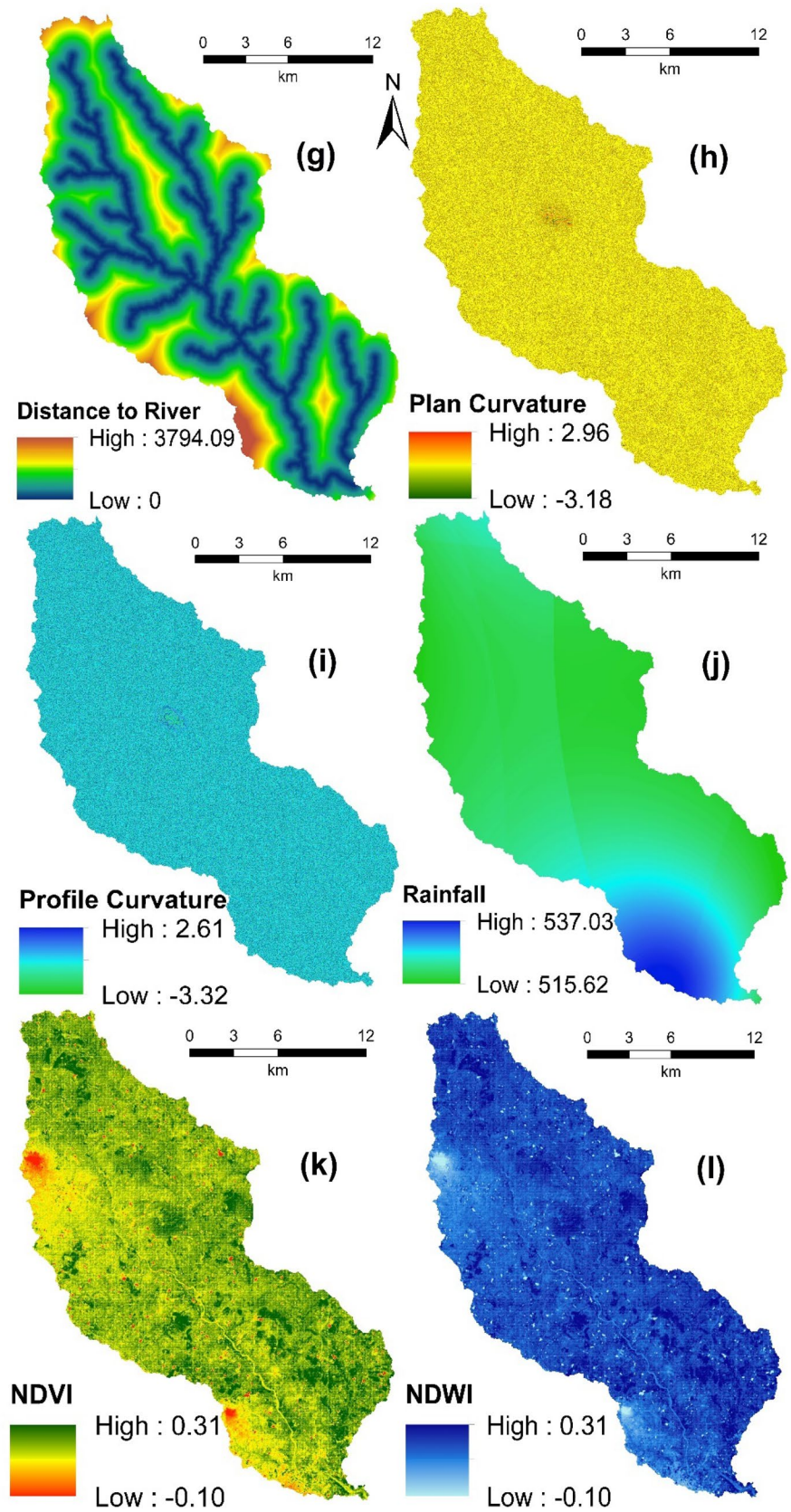
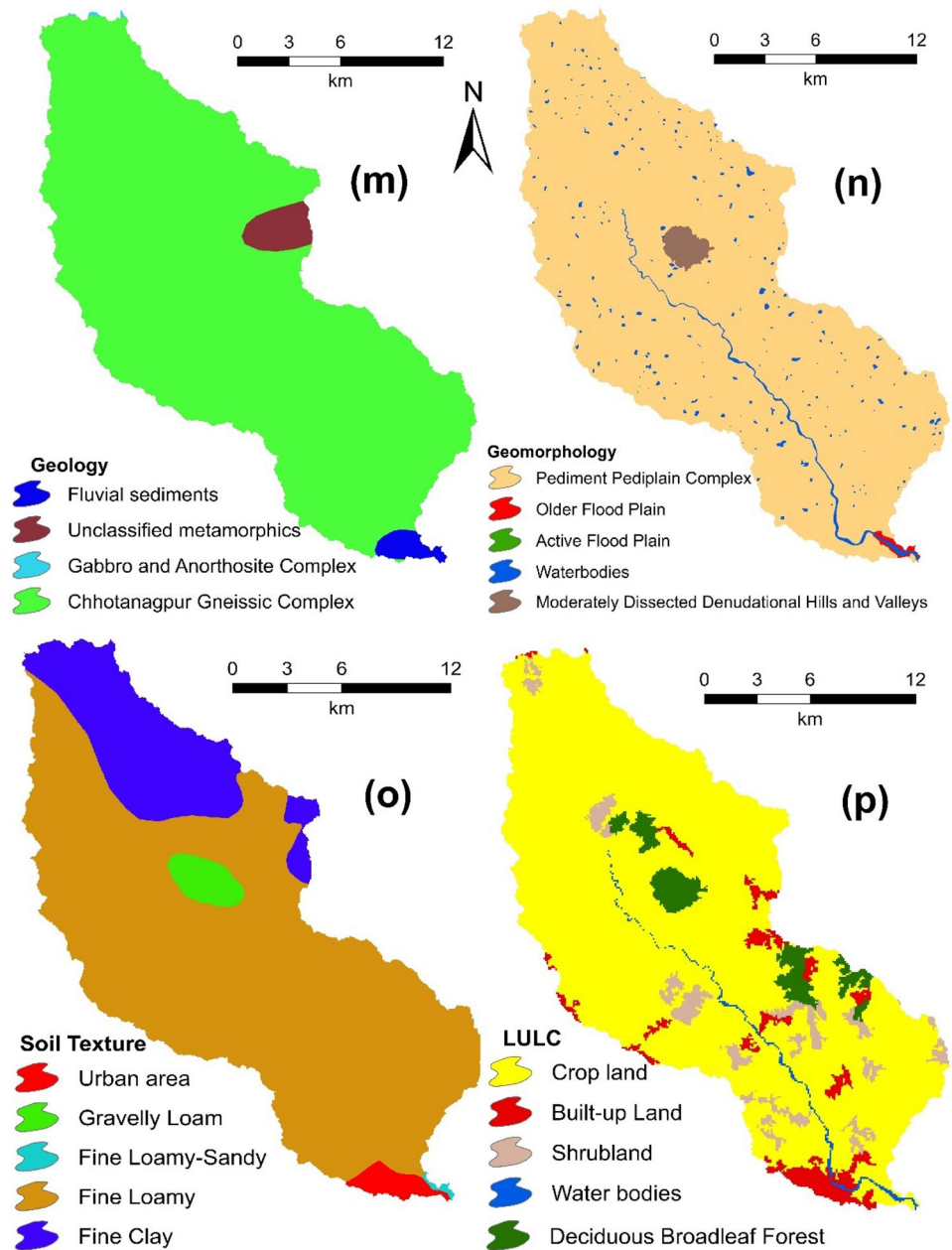


Fig. 3 (continued)



NDVI represents the vegetation status of an area. Vegetation characteristics are different based on their geomorphological location. Flood area’s plants are significantly different from the other plants in the surrounding region that can help to identify the flood-susceptible areas (Paul et al. 2019). In this study, NDVI (Fig. 3k) can be calculated through the following equation:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}, \tag{3}$$

where NIR refers to the near-infrared band.

NDWI is the index that indicates the amount of water in an area. This is very useful to monitor drought conditions and other studies (Gu et al. 2007) as well as control the flood events. NDWI ranges from -1 to $+1$, where high values represent the high water content and low to very low water contents in an area. In this study, NDWI ranges from -0.10 to $+0.31$ (Fig. 3l). It has been calculated from the following equation

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}, \tag{4}$$

where NIR represents the near-infrared band and SWIR represents the short wave near-infrared band.

The geological structure of an area has great control on the channel properties, runoff development, and infiltration rate of a river of an area (Malik et al. 2020). So, geology plays a significant role in determining the flood-susceptible area. In this study, four types of geological structures were identified (Fig. 3m), among them Chhotanagpur gneissic complex is the most dominant structure.

The geomorphology of any area controls the various geomorphic events, among them flash floods are very sensitive events. To make proper FFSM, geomorphology helps significantly because several geomorphic features have their influences on flood occurrence. The current study identified five geomorphic areas in general (Fig. 3n); pediment pediplain complex is the dominant one in this region.

Soil texture has a great impact on flood events which carries the major flood vulnerability indication, because it has a strong impact on holding water, infiltration rate, and subsurface flow formation. The surface runoff and intensity of flood levels are significantly influenced by the soil composition (Cosby et al. 1984). Different soil textures and hydraulic properties have different effects on runoff formation (Lovat et al. 2019). In this study, five types of soil texture (Fig. 3o) are present which have different effects on flood occurrence.

LULC has direct and indirect control on infiltration rate, surface runoff, and evapotranspiration of an area (Bui et al. 2018; Roy et al. 2020); this emerges as a broadly accepted flash flood-indicating parameter in recent times. For example, high vegetation cover increases infiltration rate and reduces runoff, decreasing the probability of flood events, whereas bare land and urbanized area are highly vulnerable to flood (Ngo et al. 2021). In recent times, most of the flood events occur due to the encroachment toward the river channel and changes in land use and land cover pattern in the riparian area. Several studies give much more importance to LULC in case of flood events. LULC change is the significant reason for flash flood occurrence in this study area (Fig. 3p); in this study, LULC map shows that the southern portion of the study area characterized a significant built-up area along the river channel that may trigger flash flood occurrence.

Multicollinearity assessment

Several flood causative factors were used to predict flash flood-susceptible areas. Multicollinearity test has been used to minimize the probable error in modeling associated with variables. If two or more variables are highly correlated with each other than the model, it decreases the accuracy of the output result (Saha et al. 2021c). Multicollinearity (MC) helps identify those factors and characterize a linear dependency between those factors (Arabameri et al. 2018;

Saha et al. 2021b). Usually, MC assessment is made with variance inflation factor (VIF) and tolerance techniques (TLT). Various studies (Khosravi et al. 2019; Band et al. 2020b; Malik et al. 2020) show that if VIF value > 10 and TLT value < 0.10, there is a multicollinearity issue. For getting better result in FFSM, VIF values should be less than 5. This MC is determined by the following equation:

$$TLT = 1 - r^2, \quad (5)$$

$$VIF = \frac{1}{TLT}. \quad (6)$$

Models used in flash flood-susceptibility assessment

Till now, several methods had been used to get an accurate result in identifying flood-susceptible areas by various researchers. In the present time, ML has become an essential approach to make a better result because of its high accuracy and precision level with large dataset. In this study, three important ML approaches, such as support vector regression (SVR), the ensemble of support vector regression with particle swarm optimization (SVR-PSO), and grasshopper optimization algorithm (SVR-GOA), have been used to estimate the exact flash flood-susceptible area.

Support vector regression (SVR)

The SVR modeling approach is rooted in statistical learning or Vapnik–Chervonenkis (VC) theory (Awad and Khanna 2015), and at the beginning it was used to solve the pattern recognition problem. In 1999, Vapnik (Vapnik 1999) promoted the SVM method to deal with function fitting problems which form the support vector regression method (Rahman and Hazim 1996). This supervised ML algorithm is characterized by the use of sparse solution, kernels, and VC control of the margin and number of support vectors (Drucker et al. 1997; Awad and Khanna 2015). SVR algorithm helps to develop the structure and control complex functions within a system (Saha et al. 2021a). Maximization of nominal margin through regression task analysis is the significant benefit of the SVR model (Li et al. 2010). It also gives us the flexibility to describe how much error is tolerable in our model and will find the hyperplane to fit the data (Sharp 2020). In the case of a complex dataset, the SVR model is mostly used and solves this dataset by developing several curved margins (Kalantar et al. 2018). In these ML algorithms, the structural risk minimization norm (SRM) plays a very significant role in identifying the relationship between the input and output variables (Saha et al. 2021a). Thus, it is necessary to calculate SRM in an SVR model and this can be calculated using the following equations.

$$y = k(z) = v\theta(z) + c, \tag{7}$$

in which input data and the resultant value are shown by $z = (z_1, z_2, \dots, z_n)$ and $y_b \in R^l$, respectively. In addition to this, $v \in R^l$ represents the weightage factor, $c \in R^l$ represents the constant number of the mathematical function, and l represents the data size in the respective model. $\theta(z)$ represents the irregular function for map of the input dataset. To define v and c , the following equation can be used and developed based on the SRM principles:

$$\begin{aligned} \text{Minimize : } & \left[\frac{1}{2} \|v\|^2 + P \sum_{b=1}^l (\zeta_b + \zeta_b^*) \right], \\ \text{Subjectto : } & \begin{cases} y_b - (v\theta(z_b) + c_b) \leq \varepsilon + \zeta_b \\ (v\theta(z_b) + c_b) - y_b \leq \varepsilon + \zeta_b^* \\ \zeta_b, \zeta_b^* \geq 0 \end{cases}, \end{aligned} \tag{8}$$

where the penalty factor is represented by P , ζ_b, ζ_b^* indicates loose variables, and ε represents the optimized performance of the model. The Lagrangian function was used to solve the optimization problem:

$$\begin{aligned} L(v, c, \zeta_b, \zeta_b^*, \beta_b, \beta_b^*, \delta_b, \delta_b^*) \\ = \frac{1}{2} \|v\|^2 + P \sum_{b=1}^l \left(\zeta_b + \zeta_b^* - \sum_{b=1}^l \beta_b (\zeta_b + \varepsilon - y_b + v\theta(z_b z) + c) \right. \\ \left. - \sum_b \beta_b^* (\zeta_b^* + \varepsilon + y_b - v\theta(z_b z) - c) - \sum_{b=1}^l (\delta_b \zeta_b + \zeta_b^* \delta_b^*) \right), \end{aligned} \tag{9}$$

in which the Lagrangian multipliers are represented by $\delta_b, \delta_b^*, \beta_b$ and β_b^* .

Subsequently, SVR can be calculated by:

$$K(z) = \sum_{b=1}^l (\beta_b - \beta_b^*) m(z, z_b) + c, \tag{10}$$

where the kernel function is expressed through $m(z, z_b) = \langle \phi(z), \phi(z_b) \rangle$.

Particle swarm optimization (PSO)

The PSO modeling approach developed by James Kennedy and Russel Eberhart in 1995 is based on the social behaviour of fish schooling or birds flocking (Kennedy and Eberhart 1995), which is a nature-inspired algorithm and used to find out the optimal solutions of nonlinear functions. This ML method was inspired by the collective/swarm motion of biological organisms (Kennedy and Eberhart 1995). The PSO model defines each optimization issue as a bird that searches particles or space (Rizeei et al. 2019). Generally, a class of unsystematic particles that is modified by PSO

is used to explore the ideal answers by using interactive techniques (Cheng et al. 2018). This PSO is a simple, efficient, and also effective optimization algorithm that helps to solve multimodal, discontinuous, and non-convex problems (Alatas et al. 2009). Although this optimization method has experienced many changes since 1995, many researchers have developed a new version of it to understand the various aspects of the algorithm (Poli et al. 2007). In analogy, PSO is an evolutionary computation method that follows a stochastic optimization model based on swarm intelligence (Rana et al. 2011) and shows a set of solutions in a search space to achieve the best solution or position. The algorithm of PSO can be described in the following way. If dataset on a particle i uses d -dimensional representation, the location is $Y_i = (y_{i1}, y_{i2}, \dots, y_{id})^T$, the respective speed is $V = (v_{i1}, v_{i2}, \dots, v_{id})^T$, and other vectors are alike, then,

$$\begin{aligned} v_{id}^{k+1} = w * v_{id}^k + c_1 * rand_1^k * (pbest_{id}^k - x_{id}^k) \\ + c_2 * rand_2^k * (gbest_{id}^k - x_{id}^k), \end{aligned} \tag{11}$$

$$y_{id}^{k+1} = y_{id}^k + v_{id}^{k+1}, \tag{12}$$

where v_{id}^k is particle i in the k th iteration of the d -dimensional speed; c_1, c_2 are the accelerating coefficients, with $c_1 = c_2 = 2$; $rand_{1,2}$ is the random number between 0 and 1; x_{id}^k is particle i in the k th iteration of the d -dimension of the present location; $pbest_d$ is the individual extreme point position where particle i is located in the d -dimensional; $gbest_d$ is the overall situation extreme point position in which the total particle group is situated in the d -dimensional. W represents primary weight value of the inertia and the values ranging from 0 to 1.4. Here, the essential reduction can make by w for the number of iterative smaller, and in the march of time w linearly decreases.

Grasshopper optimization algorithm (GOA)

Optimization is the process of finding the best value for the variable to minimize and maximize an objective function of a particular problem. Based on objectives optimization, problems are classified into two, single objective and multi-objective optimization problems (Marler and Arora 2004). For this purpose, the mathematical optimization method has been used in many research works to make optimal results (Intriligator 2002; Dunning et al. 2017) but, in recent times, nature-inspired population-based algorithms, stochastic optimization approaches have become more popular (Yang 2010). Several nature-inspired multi-solution-based meta-heuristic algorithms such as genetic algorithms (GA) (Holland 1992), particle swarm optimization (PSO) (Kennedy and Eberhart 1995), and ant colony optimization (ACO) (Dorigo et al. 2006) have been used

in different works in the past years. Among them, grasshopper optimization algorithms (GOA) are emerging as important meta-heuristic, multiple objective-based multi-resolution approaches (Saremi et al. 2017). GOA is totally dependent on the swarm nature of grasshoppers, which gave the most accurate and competitive results than other nature-inspired optimization methods (Luo et al. 2018). This method has been applied in various fields due to its efficiency and simple implementation (Łukasik et al. 2017; El-Fergany 2017; Mirjalili et al. 2018a; Tharwat et al. 2018). The mathematical equations proposed for this model are as follows. In this model, the position of the i -th grasshopper is denoted as X_i and represented as:

$$X_i = S_i + G_i + A_i, \quad (13)$$

where S_i indicates social interactions, G_i represents gravity force on the i -th grasshopper and A_i indicates wind advection (Mirjalili et al. 2018b).

The above-mentioned equation may discuss about the movements of the grasshopper, and the social interactions of the grasshopper may be discussed as:

$$S_i = \sum_{\substack{j=1 \\ j \neq i}}^N s(d_{ij}) \hat{d}_{ij}, \quad (14)$$

where d_{ij} represents the distance between the i -th and j -th grasshopper, and s is the function that describes the strength of social forces and is represented as:

$$s(r) = fe^{\frac{-r}{l}} - e^{-r}, \quad (15)$$

where f represents the intensity of attraction and l is attractive length scale. The details about the algorithms of GOA can be found in Mirjalili et al. (2018b) and Arora and Anand (2019a).

As previously mentioned, all the methods have some drawbacks and shortcomings, which gave some biasness in the produced result. According to Zhang and Hong (2019), such algorithms have numerous disadvantages such as trapping at local optima and low population density which makes premature convergence. In the case of large and noisy datasets, SVR modeling approach will be underperformed. The PSO model easily falls into local optimum in high-dimensional space and also has a low convergence rate. GOA modeling approach also faced problems such as slow convergence and trapping in local optima (Luo et al. 2018). To eliminate the biases and overcome the shortcomings, in this research work, two ensemble methods such as SVR–PSO and SVR–GOA have been used along with SVR modeling approach.

Validation of the models

Any scientific study entirely depends on appropriate validation techniques; without it, the produced result does not apply in reality (Saha et al. 2021a). In this study, six important statistical techniques, such as specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV), kappa coefficient, and receiver operating characteristics curve (ROC)–AUC, kappa coefficient analysis with one graphical validation technique namely Taylor diagram (Taylor 2001), have been used to validate the predictive result of the aforementioned models and assess the multiple aspects of models' performances. The number of pixels was used in the accuracy assessment of flood and non-flood areas. The four statistical parameters such as true positive (TP), true negative (TN), false negative (FN), and false positive (FP) have been used to calculate the validation techniques. The accuracy of the aforementioned models depends on these validation technique values: if the values were higher, then the models gave better results and vice versa (Khosravi et al. 2019). The above mentioned statistical parameters used in this study were calculated through the following equations

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}, \quad (16)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (17)$$

$$\text{PPV} = \frac{\text{TP}}{\text{FP} + \text{TP}}, \quad (18)$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}. \quad (19)$$

ROC–AUC analysis was used to validate the models, which is a widely accepted technique. ROC analysis was completed by plotting on the X axis Y axis which are popularly known as sensitivity and specificity. It helps to assess the predictive ability of models which is a well-established technique for this type of estimation. The range from 0.5 to 1 represents the poor performance to excellent performance of the models. This ROC–AUC technique was calculated by the following equation:

$$\text{AUC} = \sum_{k=1}^n (X_{k+1} - X_k)(S_{k+1} - S_k - S_k/2), \quad (20)$$

where area under curve is represented by AUC, and sensitivity and specificity are represented by S_k and X_k , respectively.

In this study, kappa coefficient statistical techniques were also used to determine the validity of the produced result; the kappa statistic value helps to identify the difference between the observed phenomenon in the present and

the predicted results by the adopted modeling approaches (Viera and Garrett 2005). The kappa coefficient was calculated by the following equation (Carletta 1996):

$$k = \frac{P_o - P_e}{1 - P_e} \tag{21}$$

where kappa coefficient is represented by *k*, the observed samples by *P_o* and the predicted result by *P_e*.

Results

Multicollinearity assessment

It has been proved that MC assessment is necessary to make any susceptibility modeling. It decreases the causative bias, which helps to improve the accuracy of the study. In this study, MC assessment was adopted to identify the suitable factors for FFSM. In this assessment, 16 causative parameters were selected and the MC test was assessed using VIF and TLT statistical techniques. The outcome of this test showed that TLT value ranged from 0.37 to 0.77, whereas VIF value from 1.29 to 2.66. This test implies that the present research work is within permissible limit and there is no MC issue. The highest and lowest VIF values were found in slope (2.66) and SPI (1.29), respectively. In the same way, the highest and lowest TLT value found in SPI (0.77) and slope (0.37), respectively (Table 1).

Table 1 Variance inflation factor (VIF) and tolerance of flood causative factors

Variables	VIF	TOL
Aspect	1.59	0.62
Altitude	2.14	0.46
Slope	2.66	0.37
Plan curvature	1.47	0.68
Drainage density	2.65	0.37
Profile curvature	1.44	0.69
Geology	2.01	0.49
Geomorphology	1.87	0.53
Distance from river	1.69	0.59
LULC	1.54	0.64
NDWI	2.01	0.49
NDVI	1.43	0.69
Soil texture	1.49	0.67
Rainfall	1.47	0.68
SPI	1.29	0.77
TWI	1.7	0.58

Flash flood-susceptibility assessment

In this study, SVR, and its ensemble of SVR–PSO and SVR–GOA models were used to assess the flash flood-susceptible area. All three models classified the river basin area into five zones, those are very low, low, moderate, high, and very high, to better understand spatial variation of flood-susceptible areas (Table 2).

In the flash flood-susceptible map prepared by the SVR model (Fig. 4a), the areal coverage of very high and high zone was approximately 46.23 km² (12.67%) and 36.23 km² (9.93%), which were mainly located in the middle to southern portion along the river. This flood significantly impacted the people in the riparian area who lived in close proximity to the river channel. The rest of the area is under moderate, low and very low flood-prone regions and the spatial coverage was approximately 37.43 km²(10.26%), 153.36 km²(42.03%) and 91.62 km²(25.11%), respectively.

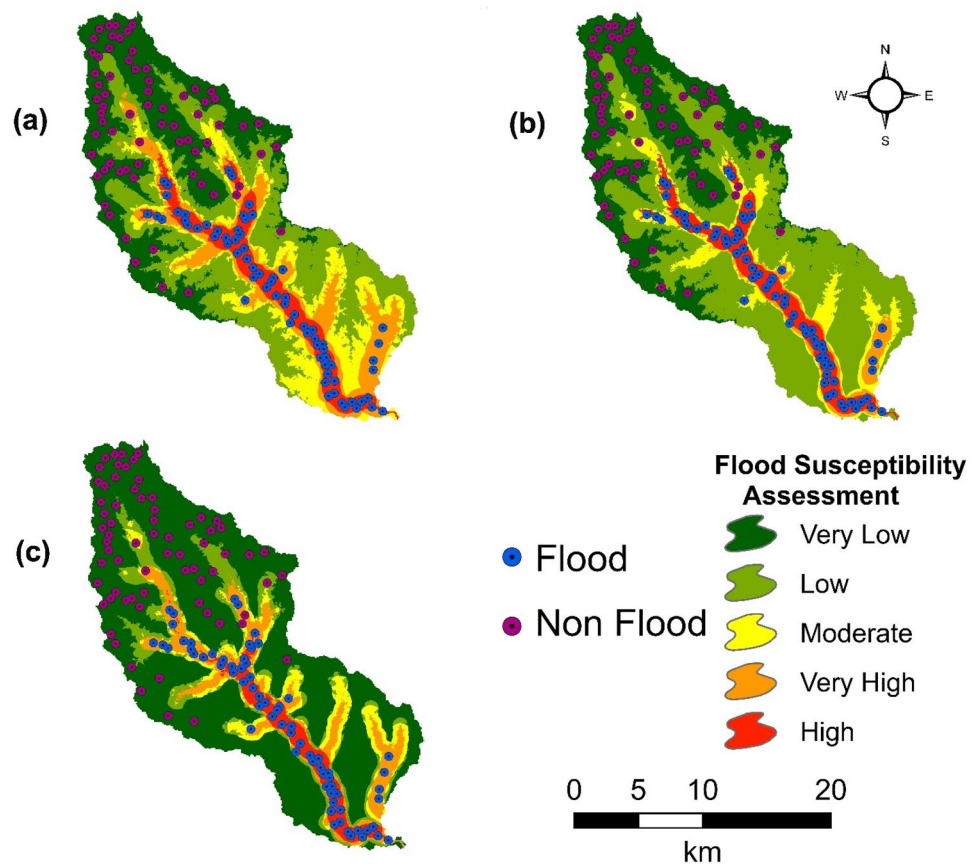
In the ensemble of the SVR–PSO model (Fig. 4b), the very high and high susceptible areas are situated in the middle to the southern stretch of the river channel, which is mostly adjacent to the river channel. The spatial coverage of very high and high zones was approximately 47.21 km²(12.94%) and 28.79 km²(7.89%), respectively, of the river basin. On the other hand, most of the area contains a low susceptible zone, which is mostly situated in the northern portion of the river basin. Spatial coverage of moderate, low, and the very low zone was approximately 27.69 km²(7.59%), 164.97 km²(45.21%) and 96.22 km² (26.37%), respectively, of the basin.

In the SVR–GOA model (Fig. 4c), the spatial coverage of the very high and high zones is 40.10 km²(10.99%) and 25.94 km²(7.11%), respectively. This model showed that a very high amount of area was under a very low flood-susceptible zone, about approximately 258.38 km²(70.81%), and the rest of the area was under low and moderate zones. The spatial coverage was approximately 19.26 km²(5.28%) and 21.20 km²(5.81%), respectively. These results imply that the occurrence of flash flood susceptibility in the downstream region of the northeast portion (near to the river) of the study region is more probable than in the upstream part of the northwest portion.

Table 2 Spatial coverage of three models in study area

Models	Flood susceptibility (area in %)				
	Very Low	Low	Moderate	High	Very High
SVR	25.11	42.03	10.26	9.93	12.67
SVR–PSO	26.37	45.21	7.59	7.89	12.94
SVR–GAO	70.81	5.28	5.81	7.11	10.99

Fig. 4 Flash flood-susceptibility mapping by using ML algorithms; **a** support vector regression SVR, **b** ensemble of SVR–PSO, **c** ensemble of SVR–GOA



Model validation

All of the models used in this study needed to be validated. The validation was done using the AUC value from the ROC curve and various statistical indices such as specificity, sensitivity, PPV, NPV, kappa coefficient and one graphical model including Taylor diagram. As mentioned, all the models play a significant role in analyzing and determining the accuracy of the adopted model's result. In this present study, validation is necessary because all the data collected from different sources were remotely sensed, so there was some possibility of error. Here, all the predicted results of different models were obtained using training and validating datasets. In this present research, AUC analysis of the aforesaid ML models is presented in Fig. 5a, b. Based on the result, it was clear that all the models had high accuracy, but the SVR–GOA model was the most suitable modeling approach among other models for this research work because of its AUC values, which were 0.951 and 0.938 for training and validation, respectively, followed by SVR–PSO (training 0.948, validating 0.924) and SVR (training 0.911, validating 0.901). Apart from ROC, all the statistical indices showed (Table 3) that SVR–GOA modeling approach has high predictive capacity in FFSM. The kappa value of SVR–GOA (training = 0.82 and validation = 0.79) also state

that this ensemble approach is the most optimal, followed by SVR–PSO and standalone PSO. From the Taylor diagram, we can easily determine the efficiency of all the predictive models and, from this, it is established that the SVR–GOA model is most optimal than SVR single alone model and ensemble of SVR and PSO (Fig. 6). The effect plots are probability plots that demonstrate the outcomes of a variable or relationship. It is essential to be using them to categorize significant results. The spots that appear to be outliers are the true critical influences, not those that fall within the conventional probability plot line. The probability plot clearly shows that the SVR–GOA is the most optimum model when evaluating the likelihood ratio (Fig. 7). Additionally, the RMSE variation graph is also presented in Fig. 8 based on individual learning algorithms, i.e., SVR, SVR–PSO and SVR–GOA.

Factor importance analysis

All the adopted models have their principle to estimate the importance of each causative factor. In the present study, all predicted models were used to estimate the variable importance in FFSM. In the SVR model, the most significant variables in forecasting flash flood susceptibility were rainfall (100), altitude (86.22), geomorphology (63.39),

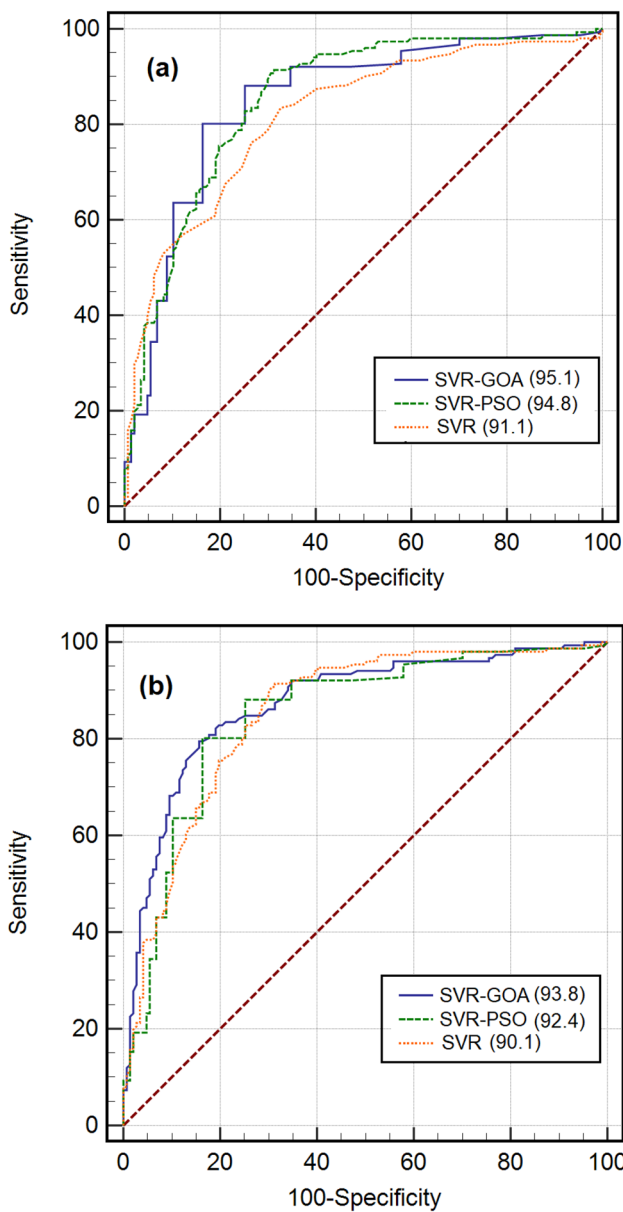


Fig. 5 Model validation by ROC curve analysis; **a** training dataset, **b** validating dataset

slope (52.21), NDWI (41.49), and TWI (32.45), while in PSO modeling the controlling factors for FFMS were rainfall (100), NDWI (44.83), geology (34.11), altitude (31.67), NDVI (29.61), and slope (28.09). In the case of the GOA model, the most important variables were rainfall (100), altitude (74.69), slope (55.99), geomorphology (41.11), NDWI (39.68), and distance from the river (32.48). Besides these variables, other variables adopted in this study by all predicting models have their moderate to low importance (Table 4).

Discussion

Flash floods are one of the most dangerous and complex natural disasters due to their short time occurrence, high-speed water runoff, and huge sediment transportation that can result in sudden destruction in properties as well as human lives. However, its complete prevention is impossible. Because of this, it is critical to develop flood prediction and mitigation strategies to save lives and reduce the socioeconomic impacts of flood events, which pose several challenges to the local authorities. Several attempts have been done by different researchers to develop a proper mitigation strategy; among them, FFMS is one of the crucial flood mitigation strategies by which flood-prone areas can be easily identified and proper structural and non-structural procedures can be adopted to minimize the effects of flooding. Several methods and modeling approaches have been used to identify flood areas, but it is also a crucial task to find out the appropriate methods that have more liability and predictivity. The prime objective of our study was to explore the suitability of employing ensemble models (SVR-PSO and SVR-GOA) in flash flood-susceptibility mapping. A significant number of studies have performed spatial modeling using ensemble models (Al-Abadi 2018; Arora and Anand 2019b; Band et al. 2020b; Roy et al. 2020; Pham et al. 2021a; Towfiqul Islam et al. 2021). In the recent era, ML algorithms and artificial intelligence have attracted substantial interest, especially for evaluating environmental hazards, because of their high predictive

Table 3 Predicting capability of three models by six adopted statistical techniques

Models	Stage	Parameters					
		Sensitivity	Specificity	PPV	NPV	AUC	Kappa
SVR	Training	0.93	0.80	0.82	0.91	0.911	0.76
	Validation	0.89	0.77	0.77	0.89	0.91	0.74
SVR-PSO	Training	0.87	0.84	0.85	0.87	0.901	0.79
	Validation	0.86	0.84	0.82	0.87	0.924	0.75
SVR-GOA	Training	0.99	0.97	0.97	0.99	0.951	0.82
	Validation	0.91	0.87	0.86	0.91	0.938	0.79

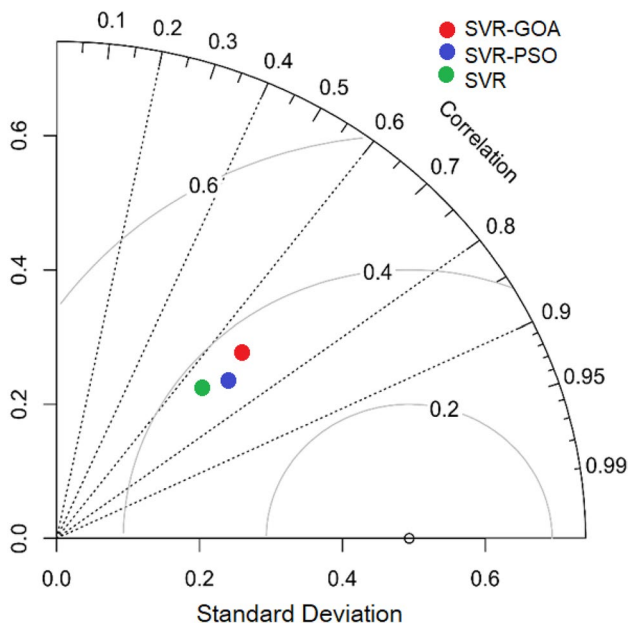


Fig. 6 graphical representation of Taylor diagram of all the applied models

accuracy and ability to work with large datasets which are less expensive (Band et al. 2020b; Fu et al. 2020). All these methods have given optimal results based on appropriate flood affecting factors in a particular region. Although substantial improvement has been made in FFSM, undoubtedly, advancement is required to improve flash flood-susceptibility mapping performance. These ML

algorithms have comparable model accuracy with several methods to model flood likelihood and differentiate the relationship between environmental effect and flood incidence (Goetz et al. 2015). A thorough investigation of this differentiation is essential to choose a suitable algorithm for a specific field of research (Bazai et al. 2020). Wu et al. (2019) and Xiong et al. (2019) have used SVR as a powerful ML technique to identify flash flood areas in China and obtained prominent results. Sahoo et al. (2021) also used an ensemble of ANFIS–GOA modeling approaches to predict the flood in the Mahanadi River basin and found notable outcomes. In predicting effective drought index, Malik et al. (2021) employed SVR–PSO algorithms where they got significant results in prediction.

Several geological, hydrological, morphological, and topographical factors are responsible for flooding (Tehrany et al. 2015). However, only a few factors have a significant role in causing flood events in a particular region, so selecting relevant factors is essential in FFSM (Khosravi et al. 2019). For instance, Luu et al. (2021) used R model and found that land use is the most important factor, followed by geology and slope largely influencing flood occurrence in Vietnam. Khosravi et al. (2018) investigated various models and found that slope was the most crucial factor in flash flood-susceptibility mapping in Iran. In this study, the three adopted models showed that rainfall was the most influential factor in flooding in this region. High intensity of rainfall in a small duration provides a massive amount of water in the river channel, flowing from higher elevation to flat area located at the lower portion

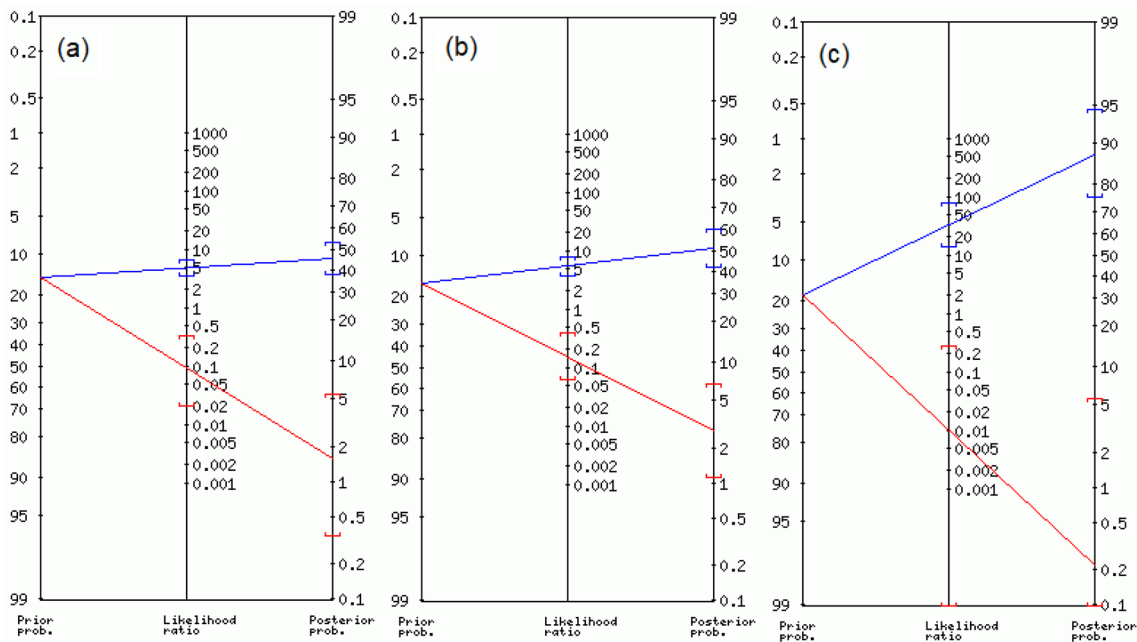


Fig. 7 Probability plot of SVR a, SVR–PSO b, and SVR–GOA c model

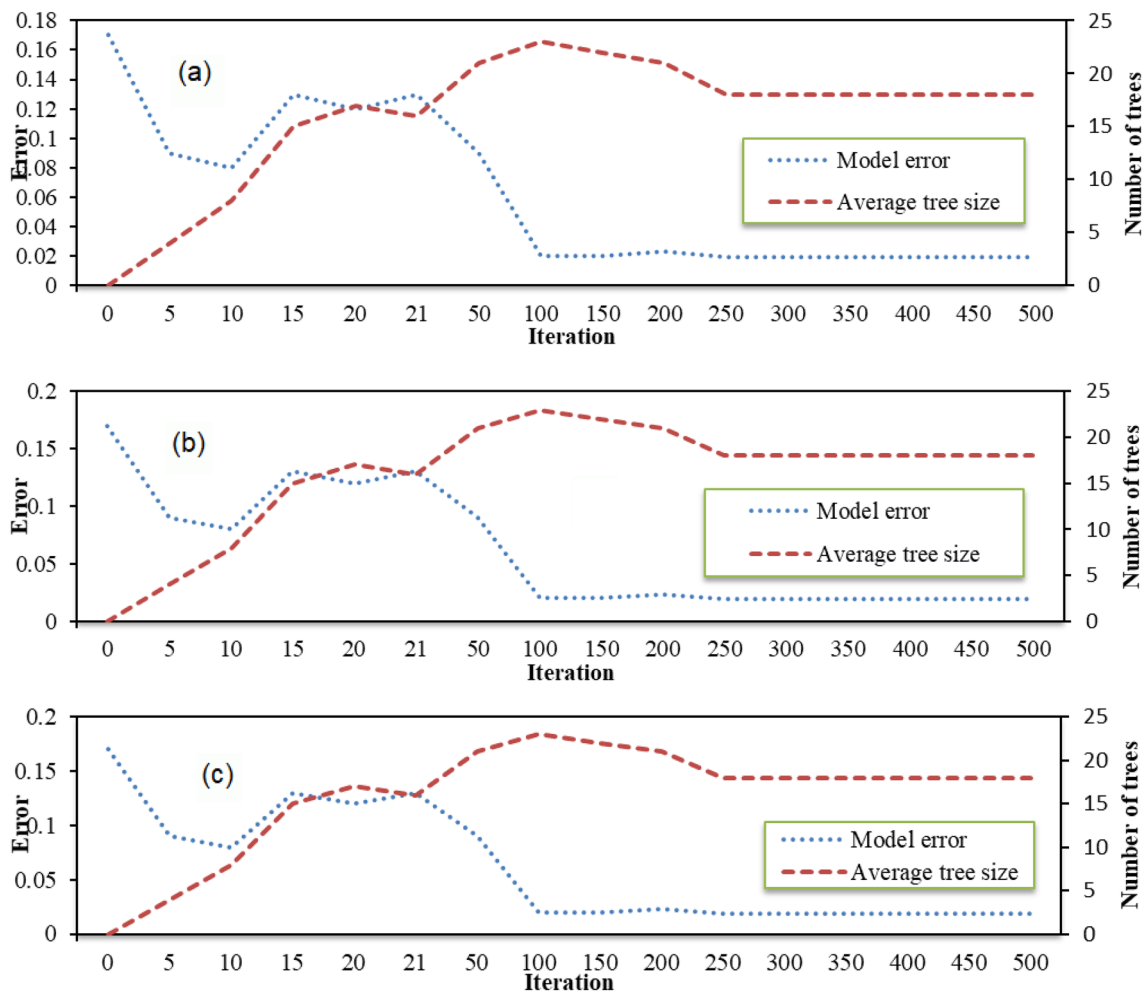


Fig. 8 Graphical representation of RMSE of three applied learning models. **a** SVR, **b** SVR–PSO, **c** SVR–GOA

Table 4 Factors importance analysis

Variables	SVR	PSO	GOA
Aspect	2.15	4.51	7.45
Altitude	86.22	31.67	74.69
Slope	52.21	28.09	55.99
Plan curvature	6.95	6.91	7.13
Drainage density	17.97	7.13	24.66
Profile curvature	8.32	6.36	5.57
Geology	11.07	34.11	27.32
Geomorphology	63.39	9.24	41.11
Distance from river	18.47	13.37	32.48
LULC	1.04	6.12	8.15
NDWI	41.49	44.83	39.68
NDVI	25.2	29.61	28.31
Soil texture	10.1	9.38	11.03
Rainfall	100	100	100
SPI	17.65	3.28	4.83
TWI	32.45	7.64	28.65

of this basin. In addition, LULC change also accelerates flood occurrence. The finding revealed that, in addition to rainfall, altitude, slope, and geomorphology were the most significant factors contributing to flash flood occurrence in the study basin.

Validation of the models mentioned above was attained by several statistical techniques such as receiving operating system (ROC) with AUC analysis, specificity, sensitivity, PPV, and NPV. The result of validation techniques reveals that the SVR–GOA model is the best approach compared to other models (SVR, SVR–PSO), which gave better result and high accuracy. This model has high predictive performance (AUC = 0.951 and 0.938, sensitivity = 0.99 and 0.91, Specificity = 0.97 and 0.87, PPV = 0.97 and 0.86, NPV = 0.99 and 0.91 in training and validation, respectively), followed by SVR–PSO and SVR model. This is verified by the fact that SVR–GOA needed less time for calculation, with less error. In contrast, the conventional SVR model, for instance, needs a higher memory, enormous dataset, and more time for computation (Kalantar et al. 2018).

Besides, conventional statistical methods need huge time, enormous datasets, and more input variables unsuitable for data-limited regions, including developing countries like India. However, the SVR–GOA algorithm showed higher accuracy compared to the other two models. It is worth mentioning that there is no previous literature concerning SVR–GOA in the FFSM. It can be said that this algorithm performed well in our case study; thus, it can predict the flash flood susceptibility very well. The proposed model can also be adapted to other natural hazards to get higher accuracy. But all the data-driven models face the most significant challenging issue like overfitting, and to eliminate this issue various techniques are being examined. All the adopted validation measures showed that the SVR–GOA ensemble models have a slight difference in training and testing dataset compared to the other two models and it can significantly overcome the overfitting problems due to the integration of two different models. Therefore, this verified ensemble model (SVR–GOA) was ready for FFSM of the entire study area. Additionally, the ensemble approach of SVR–GWO is also used globally for prediction analysis in different fields. The landslide susceptibility studies using SVR–GWO in Icheon Township, South Korea, gave an AUC result of 83% (Panahi et al. 2020), and study on landslide susceptibility in western Serbia using SVR–GWO gave an AUC result of 0.733 (Balogun et al. 2021), whereas the ensemble of SVR–GOA in this study gave an AUC result of 93.8% in the validation stage. Therefore, SVR–GOA is the most optimal in this study as compared to the previous ensemble of SVR–GWO. Furthermore, multivariate adaptive regression spline (MARS) model used in flood studies have shown AUC value of 0.93 (Mosavi et al. 2020). The application of artificial neural network (ANN) in flood studies revealed that the AUC values were 94.6% (training) and 92.0% (validation), and gave good prediction analysis (Falah et al. 2019). As compared to the MARS model, SVR–GOA gives the best performance, but in comparison with ANN, SVR–GOA gives slightly low performance in this study.

This study has several limitations. First, due to the lack of hydrodynamic information on flooding, there is a certain degree of uncertainty in the proposed method. Second, inconsistency in spatial resolution of DEM-derived causal factors may create uncertainties in the causal factors applied in the proposed way. Third, we only considered the fluvial flash flood in this research. Finally, non-flood points were randomly chosen in the study basin based on expert knowledge and previous flood records. It can generate a certain level of errors in the modeling process (Khosravi et al., 2018). Thus, future studies should focus on the model to choose reliable non-flood points to enhance the quality of the input datasets.

The result and findings of the present study would help the flood hazard managers and researchers determine the

vulnerability of flooding and in deciding ways to control and reduce the effects of flood events. Our study indicated the factors influencing flash floods. It recommends that the proposed method, a robust tool, can efficiently evaluate flood susceptibility in other basins when geo-environmental features of a specific basin and model input variables, in which hydrodynamic information may be available, are considered. Furthermore, it suggests that researchers select a suitable model to assess the flash flood susceptibility of any region. It will be noted that these modeling approaches only identify the flood-prone areas, and do not give any information about the depth or velocity of the flood. So, in the future, hydraulic models will be used to determine the flood intensity in this region.

Conclusion

In recent decades, flash flood-susceptibility assessment has been a hot topic at the national and international levels. Human encroachment on river banks and climate change are two major issues that increase flash floods in a regional context. In this study, we propose three ML algorithms including SVR and coupled ensemble algorithms (SVR–PSO and SVR–GOA) to identify the flash flood-susceptibility region of the Gandheswari River basin in West Bengal, India. In this regard, 16 environmental and topographic flood causative factors were identified, and the models were run using an empirical and reasonable approach. The multicollinearity test is also used to reduce the linear dependency between variables and improve the accuracy of adopted models. For flood inventories, flood and non-flood points were identified and used for validation. The SVR-based factor importance analysis was adopted to choose and prioritize the 16 flood causal factors for the spatial modeling. The results of SVR-derived model indicated that rainfall (100), altitude (86.22), and slope (52.21) factors mainly affected flood incidence in the study basin. In this study, it was determined that the SVR–GOA model outperformed the other two models. The AUC values of the SVR–GOA model in training and validation are 0.951 and 0.938, respectively. The spatial coverage of high and very high flash flood-susceptible regions in this model are 25.94 km^2 (7.11%) and 40.10 km^2 (10.99%), respectively, and the rest of the area was under moderate, low and very low flood-prone regions which are mostly found at the lower course of river basin near Satighat and Pathakpara regions. The primary focus of several studies was the identification of an appropriate method for FFSM. Suitable preventive measures and risk management strategies are the prime focus of FFSM. In the recent era, unplanned policy practice regarding LULC and climatic variability increased the flood events at an alarming rate (Pham et al. 2021a). As a result of this research, it is clear that the SVR–GOA

modeling approach can be used in any climatic region to determine the flash flood-susceptible zone. Overall, the proposed model is robust for flash flood-susceptibility analysis. The findings of this study can assist policymakers at the local and national levels in implementing a concrete strategy to reduce flood risk, thereby minimizing economic damage and loss of life in the study area, especially those inhabitants living near the river basin. Although the proposed approach performed very well in our case study, our study has some shortcomings. For instance, uncertainty analysis was not measured in the proposed method. Considering uncertainty with ensemble models coupled with deep learning and hydrodynamics methods should be investigated further to improve the preferences of the proposed model. Hence, risk management and flood reduction actions are indeed necessary at local administration level for preventing recurrent flood impacts on human livelihoods and regional economy. Therefore, the information regarding flood risk should be given to the inhabitants of flood-prone regions within proper time (Hegger et al. 2017). Thus, this FFMSM is an extremely important tool in the reduction of human and economic losses by implementing proper management strategies and this can help the financial sector to distribute proper compensation to the flood-affected area and to enforce the appropriate regulation regarding land use.

Funding The authors have not disclosed any funding.

Declarations

Conflict of interest All the authors declare that there is no conflict of interest.

References

- Adnan RM, Mostafa RR, Kisi O et al (2021) Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowl-Based Syst* 230:107379. <https://doi.org/10.1016/j.knosys.2021.107379>
- Aerts JCJH, Botzen WJ, Clarke KC et al (2018) Integrating human behaviour dynamics into flood disaster risk assessment. *Nat Clim Chang* 8:193–199. <https://doi.org/10.1038/s41558-018-0085-1>
- Al-Abadi AM (2018) Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. *Arab J Geosci* 11:218. <https://doi.org/10.1007/s12517-018-3584-5>
- Alatas B, Akin E, Ozer AB (2009) Chaos embedded particle swarm optimization algorithms. *Chaos, Solitons Fractals* 40:1715–1734
- Arabameri A, Rezaei K, Pourghasemi HR et al (2018) GIS-based gully erosion susceptibility mapping: a comparison among three data-driven models and AHP knowledge-based technique. *Environ Earth Sci* 77:628. <https://doi.org/10.1007/s12665-018-7808-5>
- Arabameri A, Rezaei K, Cerdà A et al (2019) A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Sci Total Environ* 660:443–458. <https://doi.org/10.1016/j.scitotenv.2019.01.021>
- Arabameri A, Arora A, Pal SC et al (2021) K-fold and state-of-the-art metaheuristic machine learning approaches for groundwater potential modelling. *Water Resour Manag* 35:1837–1869
- Arora S, Anand P (2019a) Chaotic grasshopper optimization algorithm for global optimization. *Neural Comput Appl* 31:4385–4405
- Arora S, Anand P (2019b) Chaotic grasshopper optimization algorithm for global optimization. *Neural Comput Appl* 31:4385–4405. <https://doi.org/10.1007/s00521-018-3343-2>
- Arora A, Arabameri A, Pandey M et al (2021) Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. *Sci Total Environ* 750:141565
- Awad M, Khanna R (2015) Support vector regression. In: Awad M, Khanna R (eds) *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Apress, Berkeley, CA, pp 67–80
- Balogun A-L, Rezaie F, Pham QB et al (2021) Spatial prediction of landslide susceptibility in western Serbia using hybrid support vector regression (SVR) with GWO, BAT and COA algorithms. *Geosci Front* 12:101104. <https://doi.org/10.1016/j.gsf.2020.10.009>
- Band SS, Janizadeh S, Chandra Pal S et al (2020a) Novel ensemble approach of deep learning neural network (DLNN) model and particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility. *Sensors* 20:5609
- Band SS, Janizadeh S, Chandra Pal S et al (2020b) Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sens* 12:3568. <https://doi.org/10.3390/rs12213568>
- Bazai NA, Cui P, Carling PA et al (2020) Increasing glacial lake outburst flood hazard in response to surge glaciers in the Karakoram. *Earth-Sci Rev* 212:103432
- Bubeck P, Thielen AH (2018) What helps people recover from floods? Insights from a survey among flood-affected residents in Germany. *Reg Environ Chang* 18:287–296. <https://doi.org/10.1007/s10113-017-1200-y>
- Bui DT, Panahi M, Shahabi H et al (2018) Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci Rep* 8:15364. <https://doi.org/10.1038/s41598-018-33755-7>
- Burgan Hİ, Içaga Y (2019) Flood analysis using adaptive hydraulics (ADH) model in Akarcay Basin. *Teknik Dergi* 30:9029–9051. <https://doi.org/10.18400/tekderg.416067>
- Carletta J (1996) Assessing agreement on classification tasks: the kappa statistic. *Comput Linguist* 22:249–254
- Chang L-C, Amin MZM, Yang S-N, Chang F-J (2018) Building ANN-based regional multi-step-ahead flood inundation forecast models. *Water* 10:1283. <https://doi.org/10.3390/w10091283>
- Chang TK, Talei A, Chua LHC, Alaghmand S (2019) The impact of training data sequence on the performance of neuro-fuzzy rainfall-runoff models with online learning. *Water* 11:52. <https://doi.org/10.3390/w11010052>
- Chapi K, Singh VP, Shirzadi A et al (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ Model Softw* 95:229–245. <https://doi.org/10.1016/j.envsoft.2017.06.012>
- Cheng S, Lu H, Lei X, Shi Y (2018) A quarter century of particle swarm optimization. *Complex Intell Syst* 4:227–239
- Choubin B, Moradi E, Golshan M et al (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines.

- Sci Total Environ 651:2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Chowdhuri I, Pal SC, Chakraborty R (2020) Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. *Adv Space Res* 65:1466–1489. <https://doi.org/10.1016/j.asr.2019.12.003>
- Cosby BJ, Hornberger GM, Clapp RB, Ginn TR (1984) A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resour Res* 20:682–690. <https://doi.org/10.1029/WR020i006p00682>
- Costache R, Tien Bui D (2020) Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Sci Total Environ* 712:136492. <https://doi.org/10.1016/j.scitotenv.2019.136492>
- Dodangeh E, Panahi M, Rezaie F et al (2020) Novel hybrid intelligence models for flood-susceptibility prediction: meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *J Hydrol* 590:125423
- Dorigo M, Birattari M, Stutzle T (2006) Ant colony optimization. *IEEE Comput Intell Mag* 1:28–39
- Drucker H, Burges C, Kaufman L et al (1997) Support vector regression machines. *Adv Neural Inform Process Syst* 28:779–784
- Dunning I, Huchette J, Lubin M (2017) JuMP: a modeling language for mathematical optimization. *SIAM Rev* 59:295–320
- El-Fergany AA (2017) Electrical characterisation of proton exchange membrane fuel cells stack using grasshopper optimiser. *IET Renew Power Gener* 12:9–17
- Falah F, Rahmati O, Rostami M et al (2019) Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In: Pourghasemi HR, Gokceoglu C (eds) *Spatial modeling in GIS and R for earth and environmental sciences*. Elsevier, pp 323–336
- Florinsky I (2016) Topographic surface and its characterization. Elsevier, pp 7–76
- Fu X, Pace P, Aloï G et al (2020) Topology optimization against cascading failures on wireless sensor networks using a memetic algorithm. *Comput Netw* 177:107327
- Gaume E, Livet M, Desbordes M, Villeneuve J-P (2004) Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999. *J Hydrol* 286:135–154. <https://doi.org/10.1016/j.jhydrol.2003.09.015>
- Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci* 81:1–11
- Gu Y, Brown JF, Verdin JP, Wardlow B (2007) A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophys Res Lett.* <https://doi.org/10.1029/2006GL029127>
- Guardiola-Albert C, Díez-Herrero A, Amerigo Cuervo-Arango M et al (2020) Analysing flash flood risk perception through a geostatistical approach in the village of Navaluenga, Central Spain. *J Flood Risk Manag* 13:e12590. <https://doi.org/10.1111/jfr3.12590>
- Hartnett M, Nash S (2017) High-resolution flood modeling of urban areas using MSN_flood. *Water Sci Eng* 10:175–183. <https://doi.org/10.1016/j.wse.2017.10.003>
- Hegger DLT, Mees HLP, Driessen PPJ, Runhaar HAC (2017) The roles of residents in climate adaptation: a systematic review in the case of the Netherlands. *Environ Policy Gov* 27:336–350. <https://doi.org/10.1002/eet.1766>
- Hens L, Thinh NA, Hanh TH et al (2018) Sea-level rise and resilience in Vietnam and the Asia-Pacific: a synthesis. *Vietnam J Earth Sci* 40:126–152. <https://doi.org/10.15625/0866-7187/40/2/11107>
- Holland JH (1992) Genetic algorithms. *Sci Am* 267:66–73
- Horton RE (1932) Drainage-basin characteristics. *EOS Trans Am Geophys Union* 13:350–361. <https://doi.org/10.1029/TR013i001p00350>
- Intriligator MD (2002) *Mathematical optimization and economic theory*. Society for Industrial and Applied Mathematics, Philadelphia
- Islam ARMT, Saha A, Ghose B et al (2021) Landslide susceptibility modeling in a complex mountainous region of Sikkim Himalaya using new hybrid data mining approach. *Geocarto Int* 25:1–26
- Kalantar B, Pradhan B, Naghibi SA et al (2018) Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomat Nat Haz Risk* 9:49–69
- Kanani-Sadat Y, Arabsheibani R, Karimipour F, Nasser M (2019) A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. *J Hydrol* 572:17–31. <https://doi.org/10.1016/j.jhydrol.2019.02.034>
- Kennedy J, Eberhart R (1995) Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*, pp 1942–1948 vol.4
- Khosravi K, Pourghasemi HR, Chapi K, Bahri M (2016) Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. *Environ Monit Assess* 188:656. <https://doi.org/10.1007/s10661-016-5665-9>
- Khosravi K, Pham BT, Chapi K et al (2018) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci Total Environ* 627:744–755. <https://doi.org/10.1016/j.scitotenv.2018.01.266>
- Khosravi K, Shahabi H, Pham BT et al (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J Hydrol* 573:311–323. <https://doi.org/10.1016/j.jhydrol.2019.03.073>
- Li D, Simske S, Li D, Simske S (2010) Example based single-frame image super-resolution by support vector regression. *J Pattern Recognit Res.* <https://doi.org/10.13176/11.253>
- Lovat A, Vincendon B, Ducrocq V (2019) Assessing the impact of resolution and soil datasets on flash-flood modelling. *Hydrol Earth Syst Sci* 23:1801–1818. <https://doi.org/10.5194/hess-23-1801-2019>
- Łukaszik S, Kowalski PA, Charytanowicz M, Kulczycki P (2017) Data clustering with grasshopper optimization algorithm. In: *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp 71–74
- Luo J, Chen H, Zhang Q et al (2018) An improved grasshopper optimization algorithm with application to financial stress prediction. *Appl Math Model* 64:654–668
- Luu C, Von Meding J, Kanjanabootra S (2018) Assessing flood hazard using flood marks and analytic hierarchy process approach: a case study for the 2013 flood event in Quang Nam, Vietnam. *Nat Hazards* 90:1031–1050. <https://doi.org/10.1007/s11069-017-3083-0>
- Luu C, Pham BT, Phong TV et al (2021) GIS-based ensemble computational models for flood susceptibility prediction in the Quang Binh Province, Vietnam. *J Hydrol* 599:126500. <https://doi.org/10.1016/j.jhydrol.2021.126500>
- Malik S, Pal SC (2021) Potential flood frequency analysis and susceptibility mapping using CMIP5 of MIROC5 and HEC-RAS model: a case study of lower Dwarkeswar River, Eastern India. *SN Appl Sci* 3:31. <https://doi.org/10.1007/s42452-020-04104-z>
- Malik S, Chandra Pal S, Chowdhuri I et al (2020) Prediction of highly flood prone areas by GIS based heuristic and statistical model in a monsoon dominated region of Bengal Basin. *Remote Sens Appl: Soc Environ* 19:100343. <https://doi.org/10.1016/j.rsase.2020.100343>
- Malik A, Tikhamarine Y, Sammen SS et al (2021) Prediction of meteorological drought by using hybrid support vector

- regression optimized with HHO versus PSO algorithms. *Environ Sci Pollut Res* 28:39139–39158. <https://doi.org/10.1007/s11356-021-13445-0>
- Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. *Struct Multidisc Optim* 26:369–395
- Mehrabi M, Pradhan B, Moayedi H, Alamri A (2020) Optimizing an adaptive neuro-fuzzy inference system for spatial prediction of landslide susceptibility using four state-of-the-art metaheuristic techniques. *Sensors* 20:1723
- Mirjalili SZ, Mirjalili S, Saremi S et al (2018a) Grasshopper optimization algorithm for multi-objective optimization problems. *Appl Intell* 48:805–820
- Mirjalili SZ, Mirjalili S, Saremi S et al (2018b) Grasshopper optimization algorithm for multi-objective optimization problems. *Appl Intell*. <https://doi.org/10.1007/s10489-017-1019-8>
- Mondal B, Mistri D (2015) Analysis of hydrological inferences through morphometric analysis: a remote sensing-GIS based study of Gandheswari River Basin in Bankura District, West Bengal. *Int J Hum Soc Sci Stud* 2(4):68–80
- Mosavi A, Golshan M, Janzadeh S et al (2020) Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. *Geocarto Int*. <https://doi.org/10.1080/10106049.2020.1829101>
- Ngo P-TT, Pham TD, Nhu V-H et al (2021) A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. *J Hydrol* 596:125682. <https://doi.org/10.1016/j.jhydrol.2020.125682>
- Pal I, Tularug P, Jana SK, Pal DK (2018) Risk assessment and reduction measures in landslide and flash flood-prone areas: a case of southern Thailand (Nakhon si Thammarat province). *Integrating disaster science and management*. Elsevier, pp 295–308
- Panahi M, Gayen A, Pourghasemi HR et al (2020) Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Sci Total Environ* 741:139937. <https://doi.org/10.1016/j.scitotenv.2020.139937>
- Panahi M, Rahmati O, Rezaie F et al (2022) Application of the group method of data handling (GMDH) approach for landslide susceptibility zonation using readily available spatial covariates. *CATENA* 208:105779
- Paul GC, Saha S, Hembram TK (2019) Application of the GIS-based probabilistic models for mapping the flood susceptibility in Bansloi sub-basin of Ganga-Bhagirathi River and their comparison. *Remote Sens Earth Syst Sci* 2:120–146. <https://doi.org/10.1007/s41976-019-00018-6>
- Peduzzi P (2017) Prioritizing protection? *Nature Clim Chang* 7:625–626. <https://doi.org/10.1038/nclimate3362>
- Pham BT, Jaafari A, Phong TV et al (2021a) Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. *Geosci Front* 12:101105. <https://doi.org/10.1016/j.gsf.2020.11.003>
- Pham BT, Luu C, Phong TV et al (2021b) Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *J Hydrol* 592:125615. <https://doi.org/10.1016/j.jhydrol.2020.125615>
- Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. *Swarm Intell* 1:33–57
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51:350–365. <https://doi.org/10.1016/j.cageo.2012.08.023>
- Ragab R, Bromley J, Rosier P et al (2003) Experimental study of water fluxes in a residential area: 1. Rainfall, roof runoff and evaporation: the effect of slope and aspect. *Hydrol Process* 17:2409–2422. <https://doi.org/10.1002/hyp.1250>
- Rahman S, Hazim O (1996) Load forecasting for multiple sites: development of an expert system-based technique. *Electr Power Syst Res* 39:161–169
- Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto Int* 31:42–70. <https://doi.org/10.1080/10106049.2015.1041559>
- Rana S, Jasola S, Kumar R (2011) A review on particle swarm optimization algorithms and their applications to data clustering. *Artif Intell Rev* 35:211–222
- Rizeei HM, Pradhan B, Saharkhiz MA (2019) An integrated fluvial and flash pluvial model using 2D high-resolution sub-grid and particle swarm optimization-based random forest approaches in GIS. *Complex Intell Syst* 5:283–302
- Roy P, Chandra Pal S, Chakraborty R et al (2020) Threats of climate and land use change on future flood susceptibility. *J Clean Prod* 272:122757. <https://doi.org/10.1016/j.jclepro.2020.122757>
- Saha A, Pal SC, Arabameri A et al (2021a) Flood susceptibility assessment using novel ensemble of hyperpipes and support vector regression algorithms. *Water* 13:241. <https://doi.org/10.3390/w13020241>
- Saha A, Pal SC, Arabameri A et al (2021b) Optimization modelling to establish false measures implemented with ex-situ plant species to control gully erosion in a monsoon-dominated region with novel in-situ measurements. *J Environ Manag* 287:112284. <https://doi.org/10.1016/j.jenvman.2021.112284>
- Saha A, Pal SC, Santosh M et al (2021c) Modelling multi-hazard threats to cultural heritage sites and environmental sustainability: the present and future scenarios. *J Clean Product* 320:128713
- Sahoo A, Samantaray S, Paul S (2021) Efficacy of ANFIS-GOA technique in flood prediction: a case study of Mahanadi river basin in India. *H2Open J* 4:137–156. <https://doi.org/10.2166/h2oj.2021.090>
- Saremi S, Mirjalili S, Lewis A (2017) Grasshopper optimisation algorithm: theory and application. *Adv Eng Softw* 105:30–47
- Scheuer S, Haase D, Meyer V (2011) Exploring multicriteria flood vulnerability by integrating economic, social and ecological dimensions of flood risk and coping capacity: from a starting point view towards an end point view of vulnerability. *Nat Hazards* 58:731–751. <https://doi.org/10.1007/s11069-010-9666-7>
- Sekac T, Jana SK, Pal DK (2015) A remote sensing and GIS approach to assessing multiple environmental factors leading to delineation of flood hazard risk zone in the Busu River catchment, Morobe Province, Papua New Guinea. *Melanes J Geomat Prop Stud* 1:40–55
- Sharp T (2020) An introduction to support vector regression (SVR). In: Medium. <https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2>. Accessed 2 Apr 2021
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res: Atmos* 106:7183–7192. <https://doi.org/10.1029/2000JD900719>
- Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA* 125:91–101. <https://doi.org/10.1016/j.catena.2014.10.017>
- Tharwat A, Houssein EH, Ahmed MM et al (2018) MOGOA algorithm for constrained and unconstrained multi-objective optimization problems. *Appl Intell* 48:2268–2283
- Tien Bui D, Hoang N-D, Pham T-D et al (2019) A new intelligence approach based on GIS-based multivariate adaptive regression splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. *J*

- Hydrol 575:314–326. <https://doi.org/10.1016/j.jhydrol.2019.05.046>
- Tien Bui D, Hoang N-D, Martínez-Álvarez F et al (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. *Sci Total Environ* 701:134413. <https://doi.org/10.1016/j.scitotenv.2019.134413>
- Towfiqul Islam ARM, Talukdar S, Mahato S et al (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front* 12:101075. <https://doi.org/10.1016/j.gsf.2020.09.006>
- Twele A, Cao W, Plank S, Martinis S (2016) Sentinel-1-based flood mapping: a fully automated processing chain. *Int J Remote Sens* 37:2990–3004. <https://doi.org/10.1080/01431161.2016.1192304>
- Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10:988–999
- Varu J, Sekac T, Jana SK (2020) Flood hazard micro zonation from a geomatic perspective on Vitilevu Island, Fiji. *Int J Geoinform* 16:37–47
- Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the Kappa statistic. *Fam Med* 37(5):360–363
- Wasko C, Sharma A (2017) Global assessment of flood and storm extremes with increased temperatures. *Sci Rep* 7:7945. <https://doi.org/10.1038/s41598-017-08481-1>
- Wheater HS, Jakeman AJ, Beven KJ (1993) Progress and directions in rainfall-runoff modelling
- Wu J, Liu H, Wei G et al (2019) Flash flood forecasting using support vector regression model in a small mountainous catchment. *Water* 11:1327. <https://doi.org/10.3390/w11071327>
- Xiong J, Li J, Cheng W et al (2019) A GIS-based support vector machine model for flash flood vulnerability assessment and mapping in China. *ISPRS Int J Geo Inf* 8:297. <https://doi.org/10.3390/ijgi8070297>
- Yang X-S (2010) Nature-inspired metaheuristic algorithms. Luniver Press
- Youssef AM, Pradhan B, Sefry SA (2015) Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environ Earth Sci* 75:12. <https://doi.org/10.1007/s12665-015-4830-8>
- Yuan X, Chen C, Yuan Y et al (2015) Short-term wind power prediction based on LSSVM–GSA model. *Energy Convers Manag* 101:393–401. <https://doi.org/10.1016/j.enconman.2015.05.065>
- Yuan X, Chen C, Lei X et al (2018) Monthly runoff forecasting based on LSTM–ALO model. *Stoch Env Res Risk Assess* 32:2199–2212
- Zhang Z, Hong W-C (2019) Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dyn* 98:1107–1136
- Zhao G, Pang B, Xu Z et al (2018) Mapping flood susceptibility in mountainous areas on a national scale in China. *Sci Total Environ* 615:1133–1142. <https://doi.org/10.1016/j.scitotenv.2017.10.037>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.