



# Development and assessment of a novel hybrid machine learning-based landslide susceptibility mapping model in the Darjeeling Himalayas

Abhik Saha<sup>1</sup> · Vasanta Govind Kumar Villuri<sup>1</sup> · Ashutosh Bhardwaj<sup>2</sup>

Accepted: 20 July 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

Natural disasters like landslides risk people's lives and the environment. To mitigate these hazards, scientists employ landslide susceptibility mapping that evaluates zones prone to landslides and identifies contributing factors. This study aimed to introduce ensemble landslide susceptibility models based on a statistical model (analytical hierarchy process) and a machine learning model (support vector machine) for predicting landslides in a part of the Darjeeling district, West Bengal, India. A total of 114 landslide locations were identified and randomly divided into training and validation databases, with proportions of 70% and 30%, respectively. Ten conditioning factors were considered, including rainfall, soil texture, slope, aspect, geomorphology, lithology, curvature, land use and land cover, drainage density, and lineament density. The AHP-SVM model, employing linear, polynomial, radial basis function (RBF), and sigmoid function algorithms, was applied using the training landslide and non-landslide datasets along with the spatial database of conditioning factors. Four landslide susceptibility maps were generated using this model, and their accuracy was assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC) tool with the validation dataset. Among the four ensemble methods tested (AHP-SVM\_Sigmoid, AHP-SVM\_RBF, AHP-SVM\_Polynomial, and AHP-SVM\_Linear), the AHP-SVM\_Sigmoid model demonstrated the highest degree-of-fit and prediction performance, achieving a prediction capability of 86.2%. Consequently, it is concluded that the AHP-SVM\_Sigmoid ensemble model holds promise as a novel technique for spatial landslide prediction in future studies. The findings of this study are valuable for local planning and decision-makers and can be utilized to implement landslide susceptibility methods in other regions.

**Keywords** Landslide · Machine learning · Susceptibility maps · Hybrid model · AHP-SVM · Darjeeling Himalayas

## 1 Introduction

The enormous energy, unpredictability, and instability of masses in mountainous regions of the planet, landslides commonly occur there (Roy and Saha 2019). Landslide zones are widespread in the Indian Himalayan area, including those in Jammu and Kashmir, Himachal Pradesh, Kumaon, Darjeeling, Sikkim, and the hilly states in the northeast (Bhandari 2006). A landslide is a geological phenomenon that involves the movement of rock, earth, or debris down a slope due to gravity (Achour et al. 2017; Dang et al. 2020). Landslides play a major role in geomorphological change and geological catastrophes, which are ongoing hazards to the environment and human civilizations all over the globe (Chowdhuri et al. 2022a). Due to slope instability, intense rainfall, enormous masses,

---

✉ Vasanta Govind Kumar Villuri  
vgkvilluri@iitism.ac.in

Abhik Saha  
abhik.17dr000597@me.ism.ac.in

Ashutosh Bhardwaj  
ashutosh@iirs.gov.in

<sup>1</sup> Department of Mining Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad 826004, India

<sup>2</sup> Photogrammetry and Remote Sensing Department, Indian Institute of Remote Sensing, 4, Kalidas Road, Dehradun 248001, India

and seismic activity, landslides are most vulnerable and most common in mountainous areas. Landslides often occur in the mountainous Himalayan area compared to worldwide occurrences (Froude and Petley 2018; Chakraborty et al. 2022). Several types of landslides include rock falls, rock slides, debris flows, mudslides, and creep. The type of landslide that occurs depends on the properties of the material, the slope angle, and the amount of water present. Landslides can occur naturally or as a result of human activities such as excavation, deforestation, and the construction of roads and buildings (Fan et al. 2019). Landslides can significantly impact communities, causing property damage, loss of life, and economic disruption (Kjekstad and Highland 2009; Petley 2012; Skilodimou et al. 2018). They can also cause disruption to transportation networks, water supplies, and energy infrastructure. Worldwide research states that between 1995 and 2014, there were more than 3876 landslides, causing almost 163,658 fatalities and 11,689 injuries (Haque et al. 2019). According to another source, approximately 55,000 people died as a result of landslides worldwide between 2004 and 2016 (Froude and Petley 2018). According to studies, landslides have resulted in around 66,438 fatalities and 10.8 billion dollars in economic damages throughout the globe (Chakraborty et al. 2022; Chowdhuri et al. 2022a). Over 12% of India's land surface area, according to studies by the Geological Survey of India (GSI), is very susceptible to landslides. According to earlier research, several locations in the Indian Himalayan areas are more prone to frequent landslides (Pal and Chowdhuri 2019; Saha et al. 2022a). Around 80% of landslides in India occur in the Himalayas, renowned as an area with landslide problems (Chowdhuri et al. 2022b). Most landslides in the Darjeeling Himalayan ranges happen during the rainy monsoon season (June to August) as a result of precipitation seeping through lithological fractures, which affects slope instability and landslide occurrences (Chowdhuri et al. 2022b, c). The bulk of the northeastern Himalayan landslides have occurred in the Darjeeling area of West Bengal and Sikkim (Saha et al. 2022a). 72, 127, and 667 people died in historical landslides in this region in 1899, 1950, and 1968, respectively (Mandal and Mondal 2019). A landslide has significantly affected the connection of the national roads in the northeastern Himalayas, including NH-40, NH-44, and NH-44A (Mandal and Mondal 2019). The landslide events in this area are to blame for soil saturation, excessive rainfall, fractural lithology, settlement area growth, and deforestation. The Darjeeling Himalaya experiences the highest number of landslides during the monsoon season, much as the rest of the Himalayan area. Landslides have become more frequent in the Himalayan region as a result of extensive deforestation, impromptu human settlement, expansion of agriculture in

landslide-prone areas, and intense development activities, including the building of mountain roads (Chowdhuri et al. 2022b). In order to lessen the damage, suitable management techniques should be used in landslide-prone locations. Landslide monitoring involves regularly observing slopes to detect signs of instability and potential landslides. This can be achieved through geotechnical sensors, remote sensing, and ground-based monitoring. To reduce the risk of landslides, it is essential to understand the causes and triggers of these events (Devkota et al. 2013). This includes landslide conditioning factors (LCFs) such as slope, land use and land cover (LULC), geology, rainfall, earthquakes, and human activities. Because to their destructive character and negative economic effects, landslides and related phenomena have been the subject of in-depth research. In landslide-prone locations, vulnerability maps play a crucial role in the evaluation of hazardous damage and mitigation techniques. These methods are often used in regionally or watershed scale landslide assessment and prevention (Roy and Saha 2019; Das et al. 2023). Landslide-prone locations should be identified and designated using geographic information system (GIS) technology in order to compile a geographic repository of landslide inventories.

The topographic region is divided into zones with varying degrees of risk using landslide susceptibility mapping (LSM) used by decision-makers and local officials. This technique, also known as "landslide risk zoning," is essential for managing and reducing the risks related to current and potential future landslides. Using a GIS application may improve the supervision of geographical data and increase your processing power (Merghadi et al. 2020). As a result, a plethora of quantitative approaches and applications for LSM have been created. The four main categories of LSM approaches now include statistical-based models, heuristic models, physical-based models, and machine learning (ML) modeling (Arabameri et al. 2019d; Chang et al. 2019; Pham et al. 2019; Carabella et al. 2022; Selamat et al. 2022; Saha et al. 2022b). It has been demonstrated that each of these various techniques has benefits and constraints of its own. Statistical models are suitable for vast regions with geotechnical factors, whereas physical-based models work well for small areas when there is enough data for the analysis and mapping process. These models, often used to forecast future landslides, depend on in-depth knowledge of the landslide inventory obtained from adjacent subsurface and surface research and monitoring systems (Whiteley et al. 2019). Nevertheless, for lengthy research, physical-based models need a substantial quantity of precise data to get correct results (i.e., watershed up to regional level), which falls at a major financial and computational cost. The paucity of data on topography and environmental elements has an influence on both statistical-based and knowledge-

driven models, which, over the last 40 years, have dominated the area of LSM. As a consequence, physical-based models are not yet able to be used for substantial area vulnerability zonation exercises (Guzzetti et al. 1999, 2012; Ghosh et al. 2012). Landslide risk assessments involve the analysis of potential landslide locations and the likelihood of an event occurring. This information can be used to prioritize landslide mitigation measures, such as the construction of retaining walls, the planting of vegetation, and the management of land use. The management of landslide risk requires a multi-disciplinary approach involving collaboration between geologists, engineers, planners, and emergency responders. Effective landslide risk management requires a proactive approach, including the development of land-use plans, the implementation of best-practice engineering solutions, and the preparation of emergency response plans (Aleotti and Chowdhury 1999). Landslides are a serious risk that may significantly affect civilizations. It is important to recognize the roots and triggers, carry out risk assessments, and implement effective risk management strategies to reduce the risk of landslides and minimize the influence of these effects. This approach could be troublesome since it might be difficult to objectively assess or measure a result. To better understand the procedure of the patterns of landslide and triggering approaches, a number of quantitative-based models have been developed and successfully utilized. The previous ten years' GIS advancements have mostly benefited statistical-based models (Dou et al. 2019). With the introduction of statistical-based predictive models, our understanding of the susceptibility of landslides has advanced in an astoundingly short period. After the model has been built with minimal information, the landslide conditioning components in opinion-driven models (such as the analytical hierarchy process (AHP)) are ordered and weighted based on expert judgment and knowledge (Ahmed 2015). In order to create precise susceptibility zone maps, a range of LSM from various statistical-based approaches have been utilized during the last 20 years. Many authors have used these techniques for landslide mapping, and examples include the recognition of landslides with digital elevation model (DEM) derivative conditioning factors using support vector machines (SVM) technique, convolutional neural networks (CNN) for the automated detection of landslides from imagery, satellite imagery for the detection of texture changes both before and after a landslide, and valuation of the performance of SVM, random forest (RF), Principle component analysis (PCA), ANN, and CNN models (Van Den Eeckhaut et al. 2012; Kumar et al. 2017; Yu et al. 2017; Saha et al. 2020, 2021b). Different hybrid models are developed nowadays for the assessment of landslide susceptibility mapping (Shirzadi et al. 2017; Arabameri et al. 2020b; Pham et al. 2020).

Landslide susceptibility mapping in Darjeeling results from combining multiple factors in the region. Different thematic layers are derived from various sources, such as high-resolution DEM, satellite data, and published data from different organizations to facilitate accurate delineation. The area's high population density, dynamic land-use changes, and challenging topography make it an intriguing case study for understanding the intricate interplay among topography, land use, and population growth, all contributing to the risk of landslides in the Himalayan region. This research has the potential to offer valuable insights to guide land-use planning decisions and minimize the impact of landslides on local communities. The primary focus of this study is to identify vulnerable landslide areas to ensure citizens' safety and facilitate future development in the region in a large-scale area. The research's main novelty is to develop a hybrid model AHP-SVM for LSM that uses AHP and SVM to classify areas into different susceptibility classes, which is never applied.

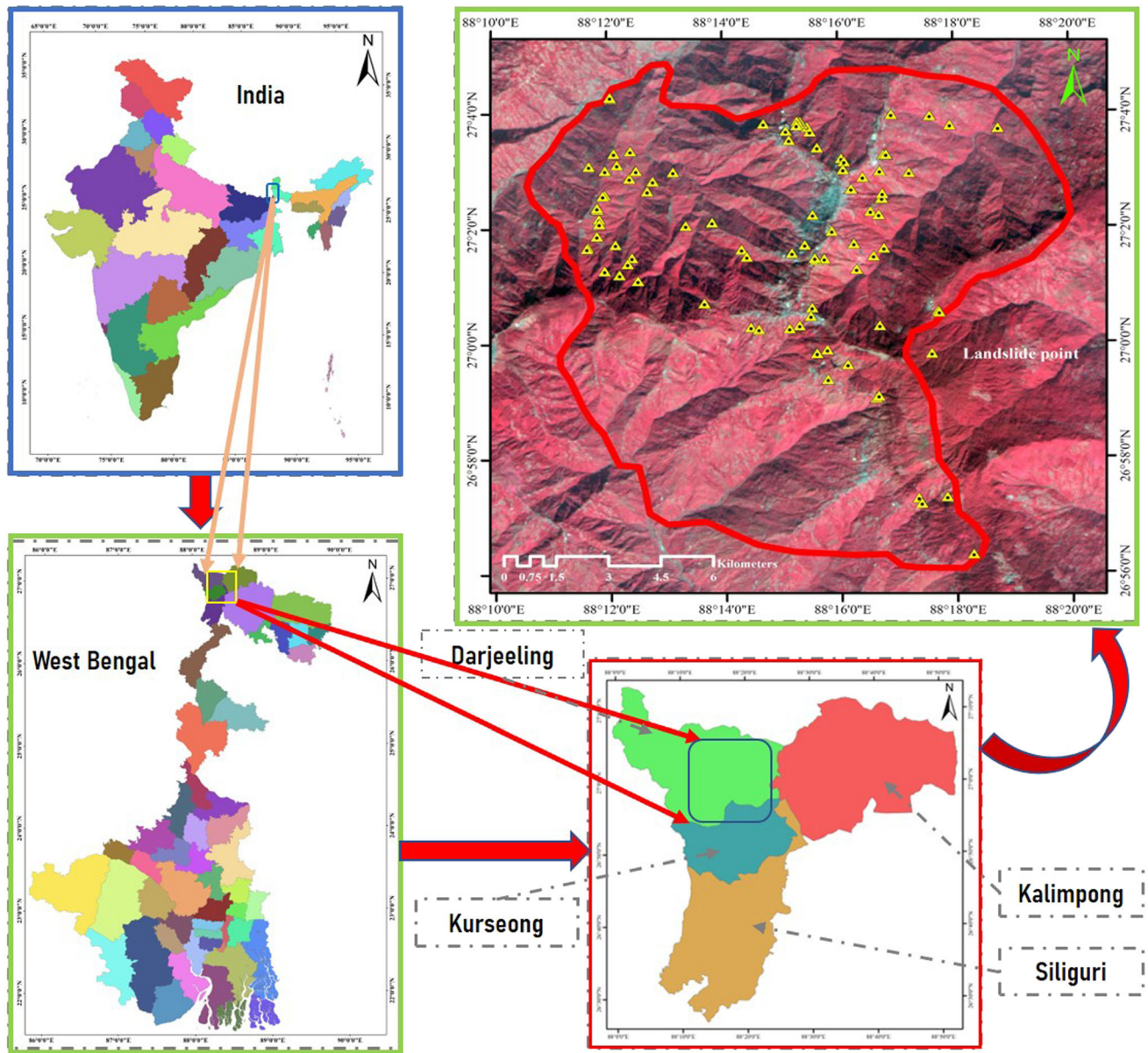
Furthermore, numerous studies employed the AHP and SVM methods to map landslide susceptibility. Still, according to the previous literature, none combined the two approaches to forecasting the probability of quantitative landslides. The receiver operating characteristic (ROC) is used to assess the models. We anticipate that site planners and local government decision-makers will utilize our research's results to lessen the risk of landslides in the region. The findings also be used by West Bengal tourism authorities to inform controlled habitation and future development strategies in areas prone to landslides.

## 2 Material and methods

### 2.1 Study area

Our research area is situated in the Darjeeling Himalayan region of India, especially in the district of Darjeeling. The study region is between latitudes  $26^{\circ}15'33''\text{N}$  and  $27^{\circ}1'22''\text{N}$  and longitudes  $88^{\circ}2'51''\text{E}$  and  $88^{\circ}5'42.6''\text{E}$  (Fig. 1). Both the plain and the mountainous topographies have distinct characteristics in this area. The research area covers a total area of  $165.92\text{ km}^2$ . The height in the region ranges from 15 to 2584 m above mean sea level (Roy and Saha 2019). The Darjeeling Himalayas are part of the Lesser and Sub-Himalayan belts. The tectonic units in the area are superposed stratigraphically in the opposite direction. Local recognition has been given to several rock bands (Kanungo et al. 2006). This area has various geomorphological features, including steep slopes and divided hills with valleys. Darjeeling's average annual temperature is approximately  $14.9\text{ }^{\circ}\text{C}$ ; however, during the winter, it dips to almost  $1\text{ }^{\circ}\text{C}$  because of the steep terrain. The





**Fig. 1** The study area represents a part of the Darjeeling district

current record low is 5 °C. Because of the area's cold climate, there is a lot of rain every year from the middle of April until the end of August. The average annual rainfall in the study region is 2074.08 mm. Significant cities like Kurseong, Darjeeling, Ghum, and Sonada exist in this area. The tea plantation and horticulture are the two features that need the most outstanding acreage. Tea, mountains, and tourism are all well-known attractions of Darjeeling. Nearly 500,000 people from India and 50,000 tourists from other countries visit Darjeeling and their surrounding areas annually. The population density is around 12,000 persons per km<sup>2</sup>.

## 2.2 Data used

Various data sources were employed in this research to provide multiple types of data (Table 1). These facts were discovered via fieldwork and historical records (data releases from the Geological Survey of India (GSI)). High-resolution satellite imagery, such as Cartosat-2D (multi-spectral) data with a resolution of 1.6 m and Advanced Land Observing Satellite-1 (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR) DEM data with a resolution of 12.5 m. Data on precipitation were gathered from the Current Research Unit (CRU). For the Indian region, soil data were taken from the ISLSCP Initiative II Data Collection (<http://www.gewex.org/is-lscp>).

**Table 1** Various data sources

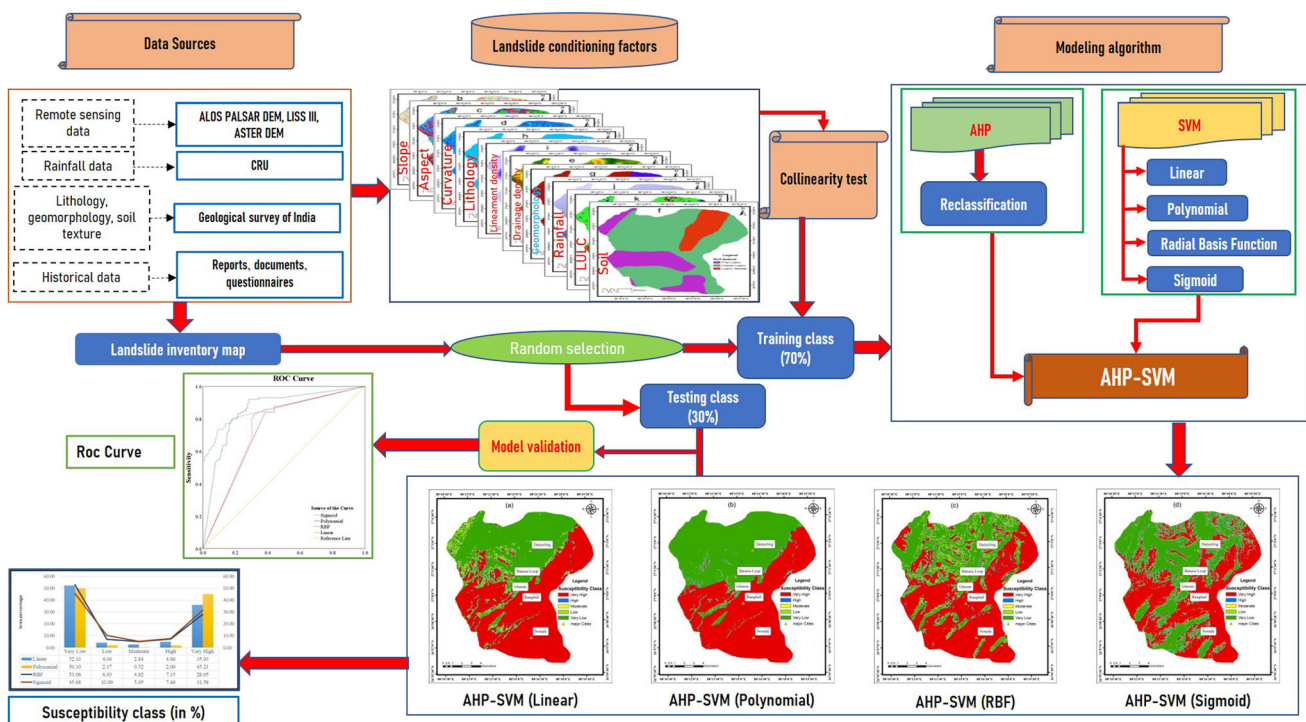
Sl.No	Input data	Resolution (spatial)	LCF maps
1	ALOS PALSAR DEM	12.5	Aspect, curvature, slope, lineament density, drainage density
2	Cartosat-2 series (MX)	1.6	LULC map
3	CRU	Numerical	Rainfall
4	Lithological and geomorphological data	–	Lithology and geomorphology map
5	Soil data	–	Soil map

html), and geomorphology and lithology data were collected from the website called BHUKOSH (<https://www.bhukosh.gsi.gov.in>).

### 2.3 Methodology

Figure 2 presents the current study’s approach. The flowchart is broken down into the following four key stages. Step 1: The first step was to create a map of the landslide inventory that has occurred in our area based on the historical data from 114 landslide locations and the same number of randomly selected training and validation points used to determine our study’s accuracy. Ten landslide conditioning factors (LCFs), including slope, aspect, curvature, lithology, lineament density, drainage density, rainfall, geomorphology, soil, and land use and land cover (LULC), were chosen for landslide susceptibility

mapping based on the results of the previous research study (Pal and Chowdhuri 2019; Arabameri et al. 2020a; Chowdhuri et al. 2020, 2022d, e) and while taking into account the topographical, climatological, and hydrological characteristics of the area. Step 2: We then used the multi-collinearity test, a statistical approach, to determine the correlation between various LCFs. The variance inflation factor (VIF) and tolerance (T) values were used to conduct the multi-collinearity test. Step 3: To create the landslide susceptibility maps (LSMs), a new AHP-SVM hybrid model was used in conjunction with ensembles of the AHP and SVM models. Step 4: To assess the models’ performance and choose the most appropriate model, LSMs were verified using the receiver operating characteristic (ROC) approach. In this research investigation, the robustness approach was also used to verify the accuracy of how each model’s maps were computed.



**Fig. 2** Graphical representation of methodology

## 2.4 Landslide inventory

It is essential to precisely identify the location and size of the landslide while developing the landslide susceptibility maps. The landslide inventory is a crucial and fundamental piece of data for every landslide zoning, including susceptibility, risk, and hazard zoning. It has to do with the location, kind, quantity, travel distance, level of activity, and timing of land sliding in a particular region (Fell et al. 2008). There are many different techniques to spot landslides. Among these are field observations, satellite images, book reviews for information on prior landslides, and aerial photography. The landslide inventory map was made using field data, visual interpretation of aerial pictures, and satellite photography (Selamat et al. 2022). The Geological Survey of India released India's comprehensive landslide inventory data. The Geological Survey of India's published data was resampled and used to create the inventory map of the research region. This research has identified 114 landslides using the published data (Fig. 1). Most of the landslides in this region that have been documented have occurred on cliffs, banks, and roadways. Landslides at cliffsides are typically brought on by erosion, terrain with steep slopes, and lack of vegetation.

## 2.5 Landslide conditioning factor

Appropriate LCFs must be considered in the scientific modeling of landslide susceptibility. The LCFs are chosen according to no set criteria. The initial literature on landslides has been reviewed in this article. The landslide conditioning elements were selected based mainly on the availability and applicability of data on a landslide occurrence. The following provides details on the 10 LCFs that were chosen for the research.

### 2.5.1 Slope

When analyzing a location's probability of experiencing a landslide, the slope is a key consideration. The study area is characterized by its hilly topography and steep slopes. That explains why there are several slopes around the basin. In the research region, the slope was divided into five classes: very high, high, moderate, low, and very low slope. Significantly few locations in the study region have very low slopes ( $< 15^\circ$ ) (Fig. 3a). The range of low slope class slope angles varies from  $15^\circ$  to  $25^\circ$ . High slopes were classified as those between  $35^\circ$  and  $45^\circ$ , and very high slopes as higher than  $45^\circ$ . Landslides usually happen in areas with high slopes; hence areas with high and very high slopes have a higher likelihood of experiencing landslides. On the other hand, landslide incidence is more or less

consistent in areas with a moderate to low slope (Tanyaş et al. 2019).

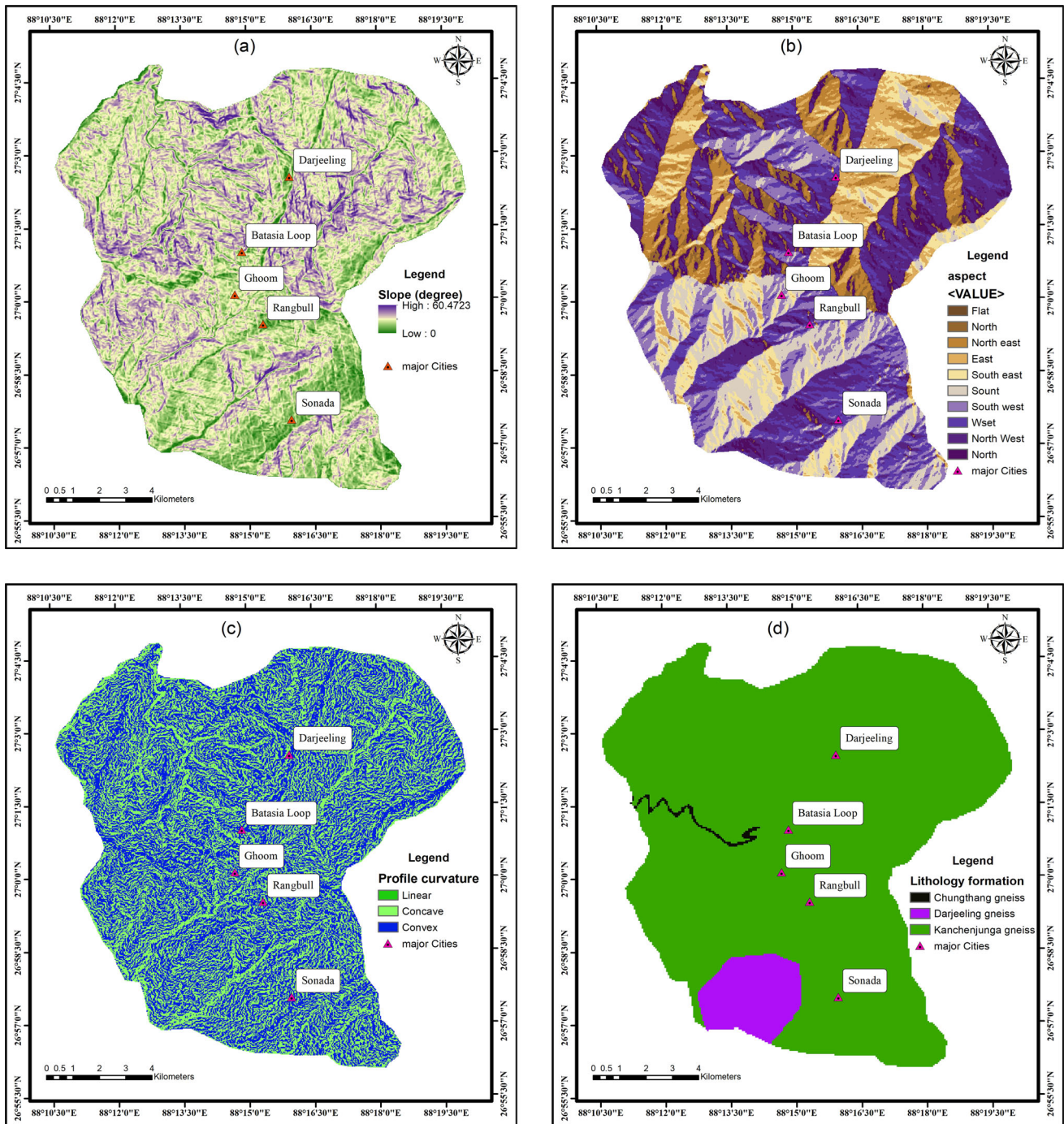
### 2.5.2 Aspect

The slope direction aspect is a crucial feature in assessing landslide susceptibility. The study area is situated in the northern hemisphere between latitudes  $26^\circ 15' 33''$  and  $27^\circ 01' 22''$ , where precipitation level and brightness of sun rays are each relatively high for slopes heading east, south, or west, respectively (Achour et al. 2017; Saha et al. 2022b). The slope that faces south is the one that is exposed to the sunlight, so evaporation happens rapidly, and it is safe than north facing slope. According to the aforementioned criteria, the slope orientation towards the north receives the lowest sun rays, which results in the highest probability of landslide, while the slope orientation towards the south receives the most sun rays and has the lowest probability of landslide. The majority of them merely experience minor effects (Mallick et al. 2018). The slope direction is shown using a 360-degree full-circle compass. A DEM was used to create the aspect map, which was then divided into nine directions as shown in Fig. 3b: (1) flat ( $-1$ ), (2) north (0 degrees to 22.5 degrees and 337.5 degrees to 360 degrees), (3) north-east (22.5 degrees to 67.5 degrees), (4) east (67.5 degrees to 112.5 degrees), (5) south-east (112.5 degrees to 157.5 degrees), (6) south (157.5 degrees to 202.5 degrees), (7) south-west (202.5 degrees to 247.5 degrees), (8) west (247.5 degrees to 292.5 degrees), and (9) north-west (292.5 degrees to 337.5 degrees). The slope facing east and south makes up a significant component of the study area. A slope angle facing west covers the lowest part of this basin, whereas the slope angle towards the north encompasses a moderate to low slope zone.

### 2.5.3 Curvature

For landslides, the curvature is a crucial topographic characteristic (Fig. 3c) (Lee and Sambath 2006; Greco et al. 2007). Landslides are less likely to occur when the curvature value lowers. The concavity (negative curvature) or convexity (positive curvature) of soil determines how much moisture it can store. The curvature value represents the topography's morphology. A positive curvature value shows an upwardly convex surface at the pixel, a negative curvature value indicates an upwardly concave surface and a flat surface is indicated by a value of 0. Convex slopes have positive curvature values, whereas concave slopes have negative curvature values (Kanwal et al. 2017). Slope segments having a higher value of either a negative or positive curvature contribute to slope saturation, drainage density, and slope instabilities. Concave slopes gather





**Fig. 3** Different landslide conditioning factors used for LSM: **a** slope; **b** aspect; **c** curvature; **d** lithology; **e** lineament density; **f** drainage density; **g** geomorphology; **h** rainfall; **i** LULC; **j** soil texture

more water, which thoroughly infuses the soil and reduces its cohesiveness (Arabameri et al. 2019a). When repeated contraction and extension cycles occur, a rock is more likely to fracture and disintegrate on a convex slope. On a convex slope, water may percolate through loosening or decomposing materials, increasing water pressure and consequential slope instability.

### 2.5.4 Lithology

The lithological characteristic of the mountain slopes significantly impacts the incidence of landslides. Lithology, which also shows the permeability and strength of the rocks, determines the characteristics of the in-situ soil that have an impact on the process of landslide (Selamat et al.

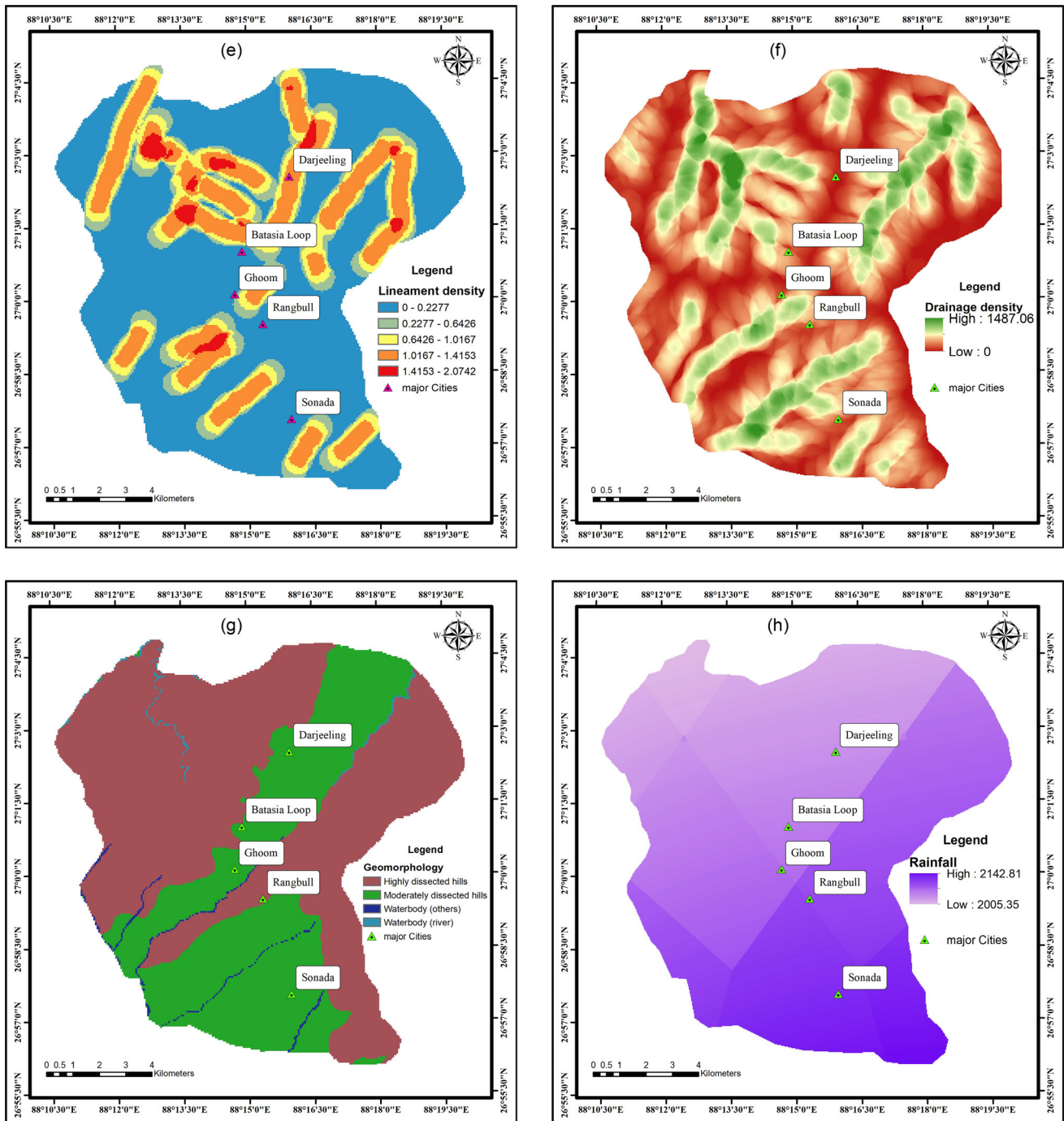


Fig. 3 continued

2022). Its composition and lithological structure determines the strength of the rock. Since they provide more significant obstruction to driving forces, stronger rocks slide less often than weaker rocks. The inverse is also applicable. To generate the lithology data, the polygons of the geological map were digitally converted into a vector layer (Fig. 3d). Three lithological types are included in this data

layer: Chungthang gneiss, Kanchenjunga gneiss, and Darjeeling gneiss.

### 2.5.5 Lineament density

In subsurface rock bodies, the visible indication of discontinuities, cracks, joints, and shear zones is called a lineament. Landslides are more likely to occur when there



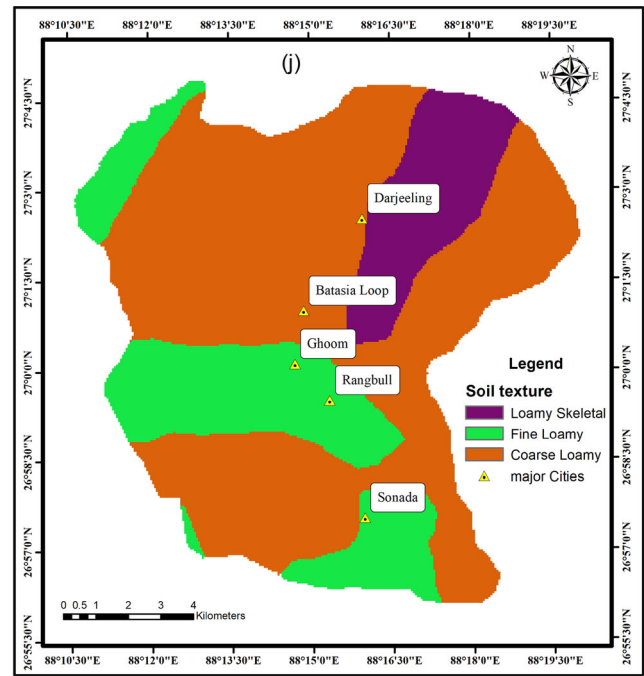
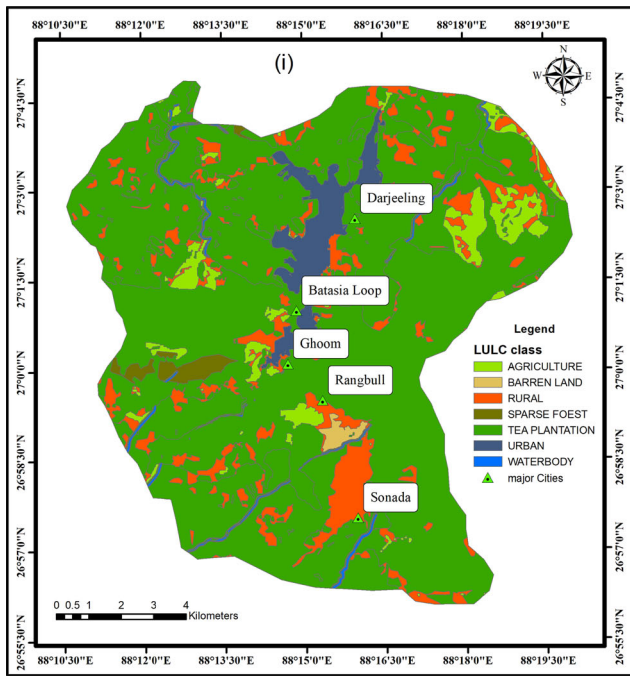


Fig. 3 continued

are lineaments that are tightly spaced. Lineaments are structural traits that show where faults, fractures, and unstable zones or planes are most prone to experience landslides. Lineaments cause slope failures by affecting the permeability of the terrain and the materials of surface structures, which have often been found to increase the chance of landslides in surrounding places (Basu and Pal 2019). Lineaments have been decoded using ALOS PALSAR DEM images. Linear stream courses that are noted for abrupt course changes, rectilinear inclinations, structural alignments of morphological characteristics, and tonal contrast were used to analyze the lineaments. Despite discovering huge lineaments, no substantial thrusts or faults have been observed in the study area. The interpreted lineaments were digitally recorded on-screen and then rasterized to provide the lineament data layer. The lineament data layer was used to construct the lineament density map (Fig. 3e), which was then separated into five classes in GIS using the natural break technique.

### 2.5.6 Drainage density

Drainage density (DD) is the number of streams per square foot in an area of a drainage basin (Fig. 3f) (Mandal and Mondal 2019). The material becomes near the base of the slope as a result of soaking up the water, which reduces the stability of the slope (Saha et al. 2022b). How much drainage density there is affects how much streams are impacted by landslides. The DD was obtained using Eq. (1) below.

$$DD = \frac{Len_d}{S_b} \tag{1}$$

where  $Len_d$  is the drainage system’s overall dimension and  $S_b$  is the drainage basin’s size. The method of Euclidean distance was used to construct drainage density maps, broken down into five categories for GIS use: 0 m–150 m, 150 m–250 m, 250 m–350 m, 350 m–450 m, and above 450 m.

### 2.5.7 Geomorphology

Geomorphology plays an essential role in determining land vulnerability. Geomorphological maps were acquired by the Indian government from the website named BHU-KOSH (<https://www.bhukosh.gsi.gov.in>). The area’s strongly divided hills and valleys play a vital role in the area’s structural genesis. It stands out due to its incline, the existence of an elevation, and the presence of a cliff that can be seen (Mandal and Mandal 2018). A unique kind of geomorphology may be found in the bottom half of the study region, which consists of a structural genesis of moderately dissected hills and valleys (Fig. 3g). The valleys in this area distinguish between terrain with high to moderate slopes and terrain with moderate to low slopes. The remaining land is covered by waterbodies (river) and waterbodies (others).

### 2.5.8 Rainfall

Rainfall is a significant initiator of landslide events in the research region (Fig. 3h) (Arabameri et al. 2020b). Rainfall causes unexpected floods and minor landslides. Water infiltrates fast due to heavy rainfall, raising the saturation level. It is an exterior trigger, and too much of it might make the slope heavier and raise pore water pressure, which could cause downhill slides. High rainfall levels change soil by increasing soil saturation, which increases the incidence of landslides. It is underlined that there is a high likelihood of a landslide occurring since severe rains have significantly separated the soil from the rock. Using Climate Research Unit (CRU) records, the method called theissen Polygon generated a rainfall distribution map. After that, it was separated into five groups using the natural breaks method, yielding the following ranges: 2005.35 mm–2032.85 mm, 2032.85 mm–2060.34 mm, 2060.34 mm–2087.83 mm, 2087.83 mm–2115.32 mm, and 2115.32 mm–2142.81 mm, since this method increases the difference between classes while minimizing the variability of data range inside a class.

### 2.5.9 Land use and land cover

The structure of a region's land use and land cover (LULC) is crucial in assessing the hazard of landslides. The LULC pattern of a region is important in the analysis of landslide danger. Land use practices in hilly areas have an enormous impact on landslide regulating elements such as cohesiveness, pore water pressure, angle of internal friction, weathering rate, and so on. Deforestation and uncontrolled building disrupt slope's natural stability, making them vulnerable to landslides in the future. Because it regulates the rate of weather and erosion, LULC is an indirect indicator of the strength of the slope. It is one of the most important elements determining the incidence of landslides in mountainous terrain. In this work, the LULC map (Fig. 3i) was produced using the CARTOSAT-2 series (MX) image. The LULC layer was digitized and rasterized with a pixel size of 30 \* 30 m. Rural areas, agricultural land, tea plantation, barren land, sparse woodland, urban area, and waterbody are all represented on the LULC map. Most of the area is covered in forestry, with agricultural land and development sections with strong road connections in the south. Due to inaccessibility, human influence is more widespread in the southern area but less so in the northern section.

### 2.5.10 Soil

The depth and kind of soil significantly influence landslide's susceptibility. The coarse-grain soil particles are

more prone to fracture due to their poor cohesion. Since the cohesiveness of soil particles is weaker than that of rock particles, the danger of a landslide rises as soil depth increases. The second non-satellite collection, the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II Data Collection, provides gridded data for 18 selected soil characteristics. These data sets have a quarter-degree resolution and are available via the Distributed Active Archive Center at Oak Ridge National Laboratory (<http://daac.ornl.gov/>). In this experiment, the soil texture (Fig. 3j) is an important consideration. There are three distinct textural groups: fine loams, skeletal loams, and coarse loams.

## 2.6 Methods

### 2.6.1 Analytical hierarchy process (AHP)

The multi-criteria decision analysis (MCDA) technique, commonly referred to as the AHP, is a structured approach for reviewing and analyzing difficult choices with a mathematical foundation. The three concepts of deconstruction, comparison, and synthesizing of the independent elements of any dependent element were the main focuses of this methodology. AHP is a qualitative and quantitative assessment approach to choosing the best option out of all available options, and it is often used to determine the relative consequence based on a pairwise comparison of the many components that are responsible for a given issue (Das et al. 2022; Saha et al. 2022b). The susceptibility-impacting factors must be included in the susceptibility map. The strata have been assigned varying weights based on their importance (Bera et al. 2019). AHP uses the pairwise comparison approach for decision-making frameworks that take a variety of criteria into account. Both subjective (qualitative) and objective (quantitative) decision-making assessments are carried out using the AHP technique (Rawat et al. 2022). The comparison matrix has an equal number of rows and columns, and one side of the matrix has the scores, while the diagonal of the matrix has a value of 1. To create a pairwise confusion matrix, the performance of each layer should be contrasted with the results of the other layers. The scale runs from 1 to 9. Each value of the pairwise comparison matrix demonstrates the significance of the two components. It was decided if feature A is given a score of 9, meaning it is more important than feature B, feature B must get a score of 1, meaning it is less important than A.  $\lambda_{EV}$  is the principal eigenvalue,  $m$  is the number of elements, and It is equal to the totals of the priority vector's individual components and the column totals.

Calculations of the consistency index (CI) and consistency ratio (CR) are necessary to verify a pairwise

comparison matrix (Chen et al. 2010; Kolat et al. 2012; Feizizadeh and Blaschke 2013). Knowing the value of the proposed random consistency index (RI) is necessary to calculate the CR (Table 2) (Saaty 2008). If the value of CR is less than 0.10, the pairwise comparison matrix may be accepted. Using the CI and CR calculation formula (Li et al. 2021),

$$CI = \frac{\lambda_{Ev} - m}{m - 1} \tag{2}$$

$$\text{Consistency Ration(CR)} = \frac{CI}{RI} \tag{3}$$

### 2.6.2 Support vector machine (SVM)

The supervised machine learning (ML) technique SVM, which is based on statistical-based learning theory, is particularly well-liked. One of the popular ML models for predicting different natural hazards is SVM, which is also used for mapping flood vulnerability and forecasting landslides. SVM is a supervised machine learning technique that provides excellent performance in a variety of fields, including the bio-informatics sector, remote sensing image categorization, pattern recognition, prediction, image processing, etc. (Saha et al. 2022c). Compared to logistic regression, which fits the data to a logistic curve and uses the existing model to estimate occurrence (Yu et al. 2010). SVM determines the best separating decision surface across classes using different training classes while attempting to maximize the margin between classes relative to the training sample (Fig. 4). The SVM technique is robust because it is model-free and data-driven, particularly for short training sets (Huang et al. 2002; Foody and Mathur 2004; Yu et al. 2010). Particularly in situations when there are more variables and fewer trainees, it offers important classification power. Support vectors (SV), a small subset of the training data points, are used as the basis for the SVM decision function (Wang et al. 2007). This method separates the construction of the hyper-plane from the training dataset. The dividing hyper-plane (HyPl) is set up between the points of two distinct classes in the real space of  $m$  coordinates ( $A_i$  parameters in vector  $A$ ). SVM reveals the highest limit of separation among the classes, and as a result, it builds a classification HyPl in the center of the highest limit.

Take into consideration a collection of linearly separable training vectors,  $A_i(i = 1, 2, \dots, m)$ . The notation identifies the two classes that make up the training vectors  $Y_i = \pm 1$ . Searching for an  $n$ -dimensional hyper-plane to separate the two classes based on their maximum gap is the aim of SVM. It may be mathematically stated as follows (Yao et al. 2008; Xu et al. 2012):

$$1/2\|V\|^2 \tag{4}$$

$$Y_i((V * A_i) + b) \geq 1 \tag{5}$$

where  $\|V\|$  is HyPl's norm and  $b$  is the scaler base.

The issue is resolved by using a Lagrangian formulation (Eq. 6). This involves adding Lagrange multipliers  $\lambda_i$  to the constraint. As a result, the objective is now to maximize with respect to  $\lambda_i$  while minimizing the Lagrangian  $L_m$  with respect to  $V$  and  $b$ . For this purpose, we used the equation below:

$$L_m = 1/2\|V\|^2 - \sum_{i=1}^m \lambda_i(Y_i((V * A_i) + b) - 1) \tag{6}$$

Four kernels—polynomial, linear, sigmoid, and radial basis function—have been employed in the current work to create the LSMs. Polynomial based kernels and radial based function kernels are mostly used kernels and are also known as Gaussian kernels. The mathematical calculation of each kernel is given below.

$$\text{Linear : } K(A_i, A_j) = A_i^T A_j \tag{7}$$

$$\text{Polynomial : } K(A_i, A_j) = (\gamma A_i^T A_j + r)^d, \gamma > 0 \tag{8}$$

$$\text{Radial basis function : } K(A_i, A_j) = (-\gamma\|A_i - A_j\|), \gamma > 0 \tag{9}$$

$$\text{Sigmoid : } K(A_i, A_j) = \tanh(\gamma A_i^T X_j + r) \tag{10}$$

$\gamma$  is the gamma function of the kernel despite of linear kernel. In the polynomial kernel,  $d$  represents the degree of a polynomial. Function  $r$  is the bias term in the sigmoid and polynomial kernel function.  $\gamma$ ,  $d$  and  $r$  are user-controlled parameters since a proper specification significantly improves the SVM solution's accuracy.

### 2.6.3 Multi-collinearity analysis

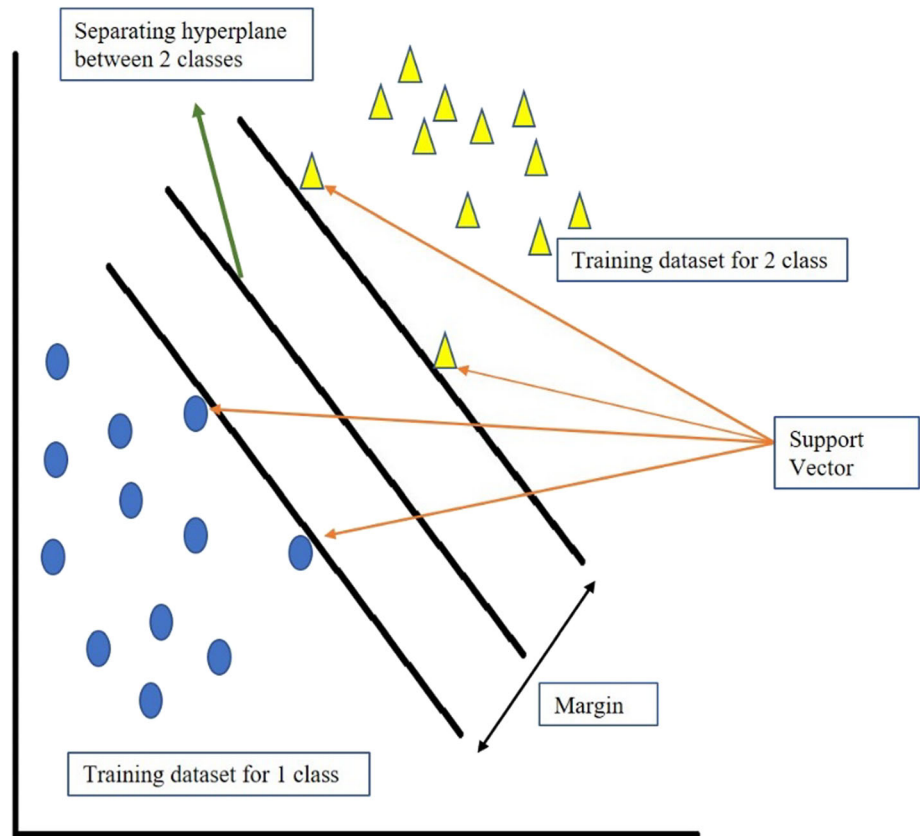
Ghosh et al. (2011) express for LSM evaluation that there are no guidelines for factor selection. Increasing the

**Table 2** Given random consistency index (RI) for CR calculation (Wind and Saaty 1980)

Sl.No	1	2	3	4	5	6	7	8	9	10	11	12	13	14
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.42	1.45	1.49	1.51	1.48	1.56	1.57



**Fig. 4** Schematic diagram of the optimal separating hyperplane



number of variables in the model's design does not constantly improve the assessment's correctness. Therefore, the first job before building the model is to examine the multi-collinearity of the chosen components (Kalantar et al. 2020). Variance inflation factor and Pearson's correlation coefficient were employed in earlier LSM research to analyze multi-collinearity (Arabameri et al. 2019a; Roy and Saha 2019; Kalantar et al. 2020; Shi et al. 2020). Multi-collinearity approaches are used in the current research to identify any collinearity among the chosen geo-environmental parameters. In the least squares regression analysis, the variance inflation factor (VIF) is used to assess the level of multi-collinearity. The exponent represents the coefficient increase calculated by multi-collinearity. The magnitude of the VIF may be used to evaluate the degree of multi-collinearity. An experimental rule states that multi-collinearity is strong if the VIF value exceeds 5 (Saha et al. 2022b). The second approach to investigating multi-collinearity is the tolerance margin of error (T). Tolerance is a broadly applicable kind of multiple correlation coefficient. Fully multicollinear variables have zero margins of error because they are utterly predictable from other independent variables. If a variable's tolerance value is 1, it is evident that it does not correlate with the other independent variables. Using the criteria VIF

and T, it was discovered that the research was multi-collinear. The formula for computing T and VIF,

$$T = 1 - R_j^2 \quad (11)$$

$$VIF = \frac{1}{\text{Tolerance}} \quad (12)$$

For  $R_j^2$ , R is the model's coefficient of determination (R-squared), where j is a descriptive variable that serves as the response variable and other explanatory variables serve as the independent variables.

#### 2.6.4 Model validation using receiver operating characteristics curve (ROC)

The efficacy of different models for landslide susceptibility may be calculated using a variety of statistically based metrics. In this work, the ROC curve of the validation model was used to evaluate the performance of the predictive model. Recently, the ROC curves approach has been used widely to evaluate the efficacy of landslide prediction (Sengupta and Nath 2022; Zhao et al. 2022). The inputs used to create the ROC curve were true positive, which on the X-axis represents a landslide that was correctly expected, and false positive, which on the Y-axis represents a landslide that was wrongly predicted. The entire efficacy of the models was quantitatively

evaluated using ROC for the side-by-side comparison. ROC model's accuracy is assigned the following grades: poor, moderate, good, excellent, and exceptional for values between 0.5 and 0.6, 0.6 and 0.7, 0.7 and 0.8, and 0.8 to 0.9 and above 0.9 (Tien Bui et al. 2016; Pham et al. 2020). ROC values may thus be used as a standard for assessing the accuracy of a prediction model.

The part of landslide sites that are appropriately categorized as landslides is the general definition of sensitivity ( $S_e$ ). While the percentage of non-landslide sites that are consistently recognized as non-landslide occurrences is the general definition of specificity ( $S_p$ ). Mokhtari and Abedian (2019; Nhu et al. 2020). Moreover, the accuracy of models assesses by the proportion of accurately detected non-landslide and landslide areas (Gautam et al. 2021). Equations used for investigation are,

$$\text{Sensitivity}(S_e) = \frac{TP}{TP + FN} \quad (13)$$

$$\text{Specificity}(S_p) = \frac{TN}{TN + FP} \quad (14)$$

$$\text{Accuracy}(A_c) = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

True negative (TN) refers to the number of correctly classified non-landslide sites, while true positive refers to the number of correctly detected landslide locations (TP). False positive (FP) and false negative (FN) points, respectively, refer to the number of landslide sites that were incorrectly recognized as non-landslide or landslide locations.

### 3 Results

#### 3.1 Multi-collinearity analysis and the correlation result

A multi-collinearity analysis is used to cross-check the validation of the implicit assumption used to choose LFCs based on the non-dependence of the components. The 10 LFCs were examined for multi-collinearity using T and VIF. When it comes to the selected independent variables, values of T less than 0.2 only weakly suggest multi-collinearity; however, T values of less than 0.1 strongly recommend it (Sujatha and Sridhar 2021; Youssef et al. 2022). The findings showed that 1.431 was the maximum VIF value, and the minimum VIF—statistic was 1.039. All VIF readings were below the theoretical threshold (5 or 10), indicating that there was no multi-collinearity among the ten conditioning variables. The maximum VIF value of the aspect was 1.431, and the minimum VIF value of 1.039 got from profile curvature, while the minimum and

maximum tolerance values were 0.743 and 0.961, respectively. As a consequence, none of the selected conditioning components exhibit multi-collinearity (Table 3).

#### 3.2 Assessment using AHP-SVM models

An important machine learning algorithm support vector machine used to determine a region's vulnerability to landslides and other natural hazards (Ma et al. 2020). The SVM classification and AHP were combined in the current investigation. The AHP-SVM classification used the landslide conditioning factors as input, including slope, aspect, drainage density, curvature, lithology, LULC, lineament density, rainfall, soil, and geomorphology. The AHP model in the GIS context provided weights to each sub-layer of the various LFCs (Table 4). The map of landslide susceptibility was then created using the weighted layers as a raster layer. The support vector machine's input data layers for ensembling with AHP were categorized as weighted layers by AHP, and a hybrid AHP-SVM approach was developed before the mapping of landslide susceptibility. The SVM classification's probabilities range from 0 to 1. The landslide susceptibility index is represented by pixels of images or conditioning variables. It has two values, ranging from 0 to 1, where 0 denotes stable conditions, and 1 indicates a high likelihood of landslides occurring. The four kinds of kernels utilized in SVM classification are the sigmoid kernel, linear kernel, radial basis function kernel, and polynomial kernel. The SVM classification utilized these functions. The landslide susceptibility maps were built in the GIS atmosphere using the combined output produced by the SVM classification. Python programming language is used for developing codes for SVM. Figure 5 shows the correlation heatmap of the AHP-SVM model developed in Google Collaboratory. A processor of Intel(R) Xeon(R) CPU E3-1245 v3 with a 3.40 GHz

**Table 3** The multi-collinearity test result of LFCs

Sl. No.	Conditioning factor	T	VIF
01	Soil	0.857	1.167
02	Curvature	0.961	1.039
03	Slope	0.876	1.147
04	Rainfall	0.796	1.251
05	Drainage density	0.922	1.085
06	Geomorphology	0.826	1.215
07	Lithology	0.944	1.060
08	Aspect	0.743	1.431
09	Lineament density	0.869	1.151
10	LULC	0.853	1.176

**Table 4** Pairwise comparison matrix of sub-factor of each landslide conditioning factor (LCF)

Factor	1	2	3	4	5	6	7	8	9	Weight
<i>Slope (in degree)</i>										
0–15	1	1/3	1/5	1/7	1/9					3.48
15–25	3	1	1/3	1/5	1/7					6.77
25–35	5	3	1	1/3	0.2					13.44
35–45	7	5	3	1	1/3					26.02
> 45	9	7	5	3	1					50.29
CI = 0.092	RI = 1.12				CR = 0.083					
<i>Aspect</i>										
Flat	1	1/3	1/5	1/7	1/8	1/9	1/7	1/3	1/2	1.88
North	3	1	1/3	1/5	1/7	1/5	1/5	1	1	3.68
North-East	5	3	1	1/3	1/4	1/7	1/3	3	3	7.43
East	7	5	3	1	1	1/2	1	1	4	14.16
South-East	8	7	4	1	1	1	3	7	9	23.92
South	9	5	7	2	1	1	5	4	8	26.96
South-West	7	5	3	1	1/3	1/5	1	5	5	13.65
West	3	1	1/3	1	1/7	1/4	1/5	1	1	5.12
North-West	2	1	1/3	1/4	1/9	1/8	1/5	1	1	3.19
CI = 0.121	RI = 1.45				CR = 0.083					
<i>Curvature</i>										
Linear	1	1/3	1/7							8.32
Concave	3	1	1/5							19.32
Convex	7	5	1							72.36
CI = 0.055	RI = 0.58				CR = 0.094					
<i>Lithology</i>										
Chungthang	1	1/3	1/7							8.99
Darjeeling Gneiss	3	1	1/5							21.34
Kanchanjunga Gneiss	7	5	1							69.67
CI = 0.047	RI = 0.58				CR = 0.082					
<i>Lineament density</i>										
1.6594–2.074	1	1/3	1/5	1/7	1/9					3.50
1.244–1.6594	3	1	1/3	1/5	1/7					6.81
0.8297–1.244	5	3	1	1/3	1/4					14.17
0.4148–0.8297	7	5	3	1	1/3					26.63
0–0.4148	9	7	4	3	1					48.89

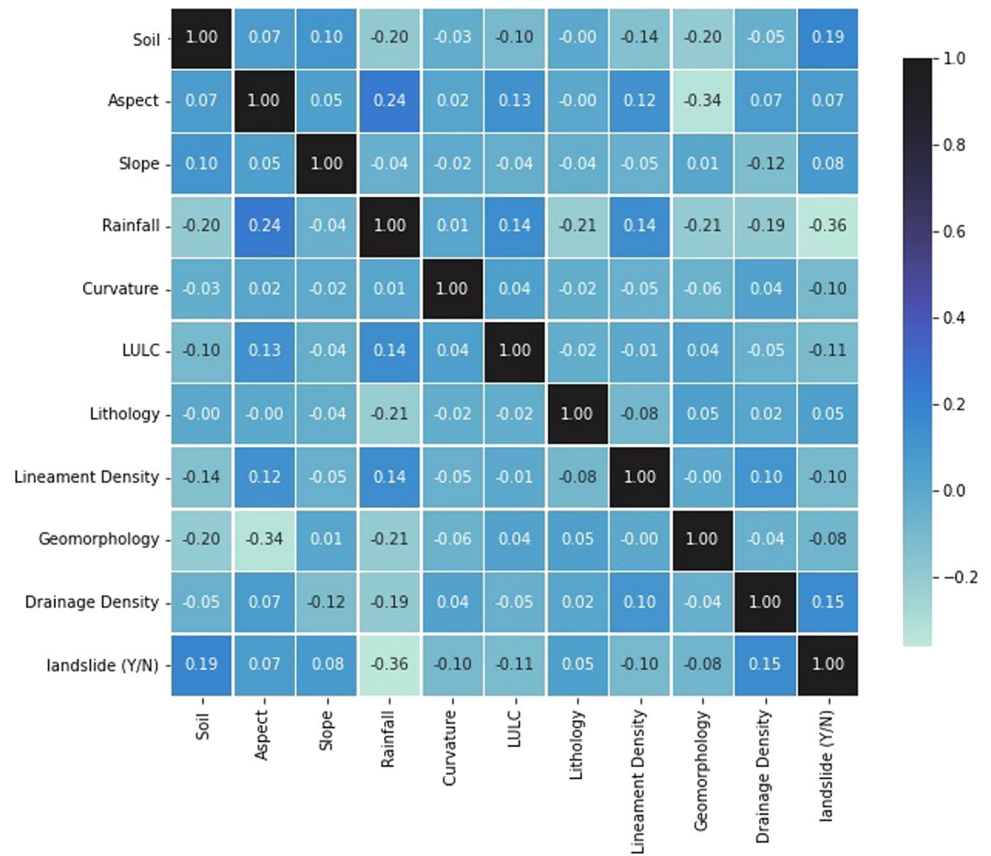


**Table 4** (continued)

Factor	1	2	3	4	5	6	7	8	9	Weight
CI = 0.084	RI = 1.12			CR = 0.075						
<i>Drainage density</i>										
> 450	1	1/ 3	1/ 5	1/ 7	1/ 9					3.48
350–450	3	1	1/ 3	1/ 5	1/ 7					6.77
250–350	5	3	1	1/ 3	1/ 5					13.44
150–250	7	5	3	1	1/ 3					26.02
0–150	9	7	5	3	1					50.29
CI = 0.092	RI = 1.12			CR = 0.083						
<i>Geomorphology</i>										
Waterbody (other)	1	1/ 3	1/ 5	1/ 7					5.69	
Waterbody (river)	3	1	1/ 3	1/ 5					12.19	
Moderate dissected hill and valley	5	3	1	1/ 5					26.33	
High-dissected hill and valley	7	5	3	1					55.79	
CI = 0.058	RI = 0.9			CR = 0.065						
<i>Rainfall (in mm)</i>										
2005.35–2032.85	1	1/ 3	1/ 5	1/ 7	1/ 9					3.64
2032.85–2060.34	3	1	1/ 3	1/ 5	1/ 6					7.31
2060.34–2087.83	5	3	1	1/ 3	1/ 5					13.92
2087.83–2115.32	7	5	3	1	1/ 3					24.35
2115.32–2142.81	9	6	5	3	1					50.77
CI = 0.078	RI = 1.12			CR = 0.070						
<i>Land use and land cover</i>										
Waterbody	1	1/ 2	1/ 3	1/ 4	1/ 5	1/ 7	1/ 9			2.76
Urban	2	1	1/ 2	1/ 3	1/ 5	1/ 7	1/ 8			3.81
Rural	3	2	1	1/ 2	1/ 3	1/ 5	1/ 7			5.91
Agricultural land	4	3	2	1	1/ 2	1/ 3	1/ 5			9.31
Tea plantation	5	5	3	2	1	1/ 2	1/ 3			15.07
Sparse forest	7	7	5	3	2	1	1/ 3			23.09
Barren land	9	8	7	5	3	3	1			40.05
CI = 0.052	RI = 1.32			CR = 0.040						
<i>Soil</i>										
Fine loamy	1	1/ 3	1/ 5							10.60
Coarse loamy	3	1	1/ 3							26.00
Loamy skeletal	5	3	1							63.40

**Table 4** (continued)

Factor	1	2	3	4	5	6	7	8	9	Weight
CI = 0.025	RI = 0.58		CR = 0.043							

**Fig. 5** Correlation heatmap of the AHP-SVM model

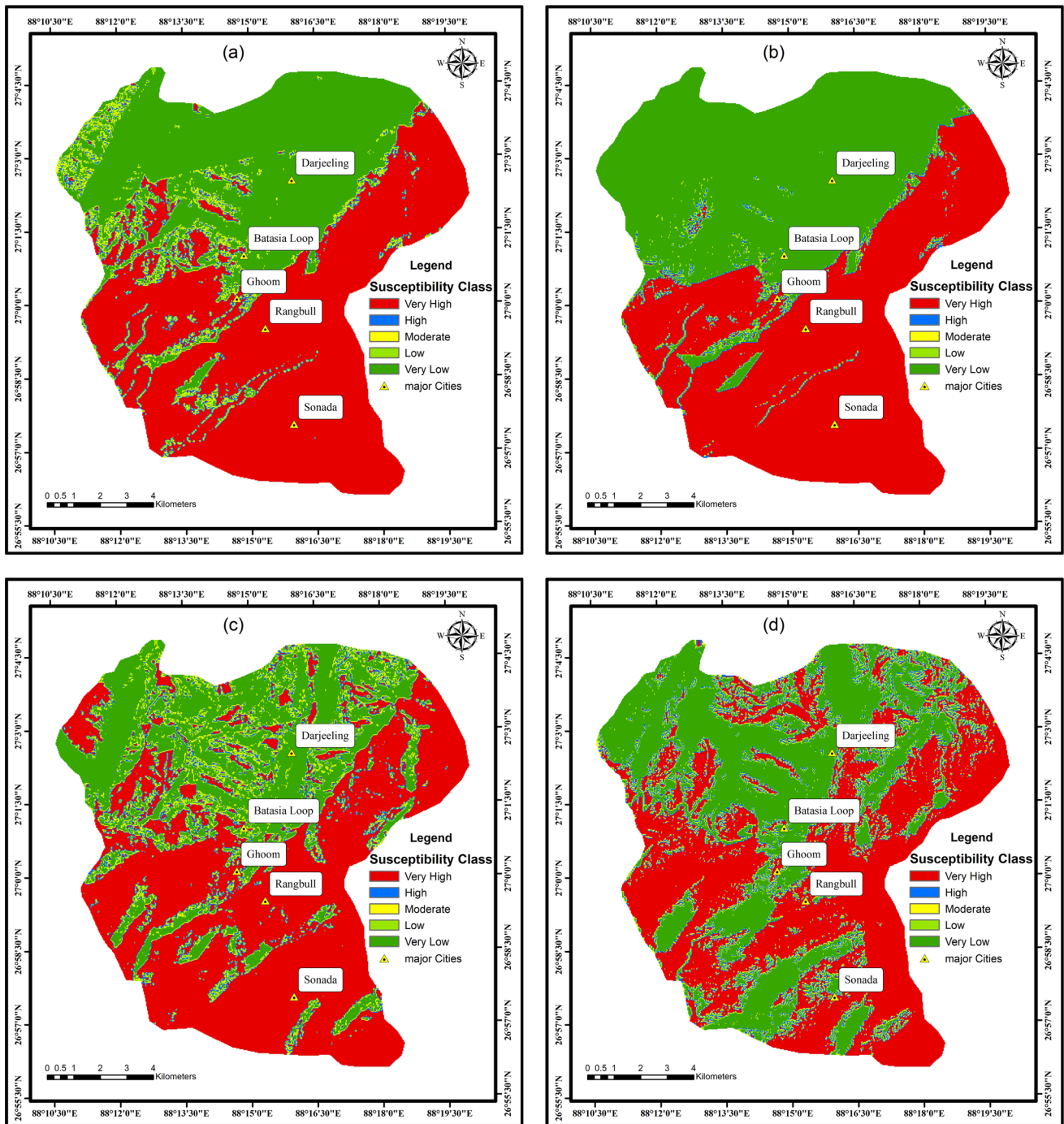
processor and 12 GB RAM is used for the performance of these models.

The four AHP-SVM hybrid models, AHP-SVM\_RBF, AHP-SVM\_Linear, AHP-SVM\_Polynomial, and AHP-SVM\_Sigmoid, were used to create the four LSMs displayed in Fig. 6a–d is produced in GIS environment after successful training, testing and successful validation of the database. These susceptibility maps were categorized into five categories: very low, low, moderate, high, and very high landslide susceptibility using the natural breaks classification technique. The five landslide susceptibility classes in the AHP-SV\_RBF ensemble map were very high, high, moderate, low, and very low. These classes, respectively, covered 8917.56 hectares (53.06%), 1165.05 hectares (6.93%), 809.91 hectares (4.82%), 1201.41 hectares (7.15%), and 4713.84 hectares (28.05%) of the region. The very high, high, moderate, low, and very low landslide susceptibility classes in the AHP-SVM\_Linear model, respectively, encompassed an area of 8795.97 hectares (52.33%), 679.5

hectares (4.04%), 476.64 hectares (2.82%), 816.3 hectares (4.86%), and 6039.36 hectares (35.93%). The very high, high, moderate, low, and very low susceptibility classes in the AHP-SVM\_Polynomial model encompassed areas of 8420.67 hectares (50.10%), 364.32 hectares (2.17%), 88.11 hectares (0.52%), 335.43 hectares (2.00%), and 7599.24 hectares (45.21%), respectively. The groups of very high, high, moderate, low, and very low landslide susceptibility covered 7677.81 hectares (45.68%), 1696.14 hectares (10.09%), 848.16 hectares (5.05%), 1277.10 hecter (7.60%), and 5308.56 hectares (31.58%), respectively, of the area in the AHP-SVM\_Sigmoid ensemble landslide susceptible map.

### 3.3 Result of model validation using ROC-AUC

The use of the ROC-AUC method was to assess how well the predictions made by the AHP-SVM kernel-based models performed. A validation dataset was used to build

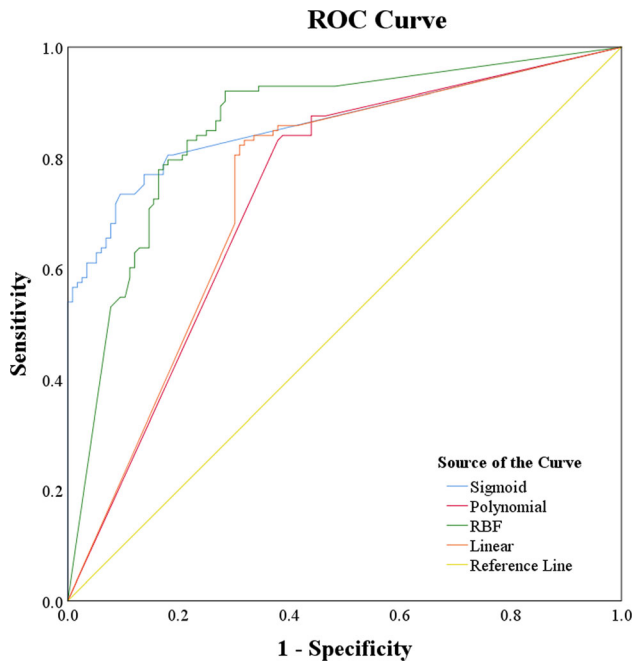


**Fig. 6** Maps of landslide susceptibility generated by various hybrid models—**a.** AHP-SVM\_Linear; **b.** AHP-SVM\_Polynomial, **c.** AHP-SVM\_RBF, **d.** AHP-SVM\_Sigmoid

the curve for prediction (Azarafza et al. 2021; Xing et al. 2021; Das et al. 2022). Thirty percent of the total landslide regions (testing datasets) were used in the current study for the model’s validation (Fig. 7). The results display that the AHP-SVM\_Sigmoid (AUC = 86.2%) based kernel performs better than the AHP-SVM\_RBF (AUC = 85.6%) based kernel, the AHP-SVM\_Linear (AUC = 74.2%)

based kernel, and the AHP-SVM\_Polynomial based kernel (AUC = 73.2%). We grouped the AUC values into various groups to provide context for the assessment values. Acceptable AUC ranges are between 0.75 and 0.8, considered fair, 0.8 and 0.9 are regarded as strong susceptibility, and AUC values above 0.9 are considered extraordinary. Our findings show the four models we used





**Fig. 7** The overall accuracy of AHP-SVM\_Linear, AHP-SVM\_Polynomial, AHP-SVM\_RBF, and AHP-SVM\_Sigmoid

had ROC-AUC values ranging from 73% to 87%, which indicates fair to strong models for all functionality and susceptibility. The results of the ROC analysis show that the AHP-SVM\_Sigmoid-based kernel functions effectively.

## 4 Discussion

Landslides have jeopardized people's safety and done significant economic harm in rugged and hilly areas. The identification of sites vulnerable to landslides in hilly terrain is one of the most crucial challenges in land use planning. LSM has been used with effectiveness during the last 30 years to address the geographic mapping of landslides (Chowdhuri et al. 2021a). While making choices in landslide-prone locations, stakeholders heavily rely on maps of landslide susceptibility. In addition to claiming human lives, landslides often devastate neighborhoods, roadways, and agricultural fields (Roccati et al. 2021). An essential tool for minimizing the risk of landslides, protecting the environmental aspect, and assisting the people of high-risk landslide susceptibility zones is the evaluation of landslide hazards using LSMs carried out in this work (van Westen et al. 2006). Landslide susceptibility evaluations are crucial in these areas because they provide a first line of defense for planners and decision-makers (Roccati et al. 2021). It is challenging to create a precise LSM that can be used to identify areas at risk for landslides (Mandal

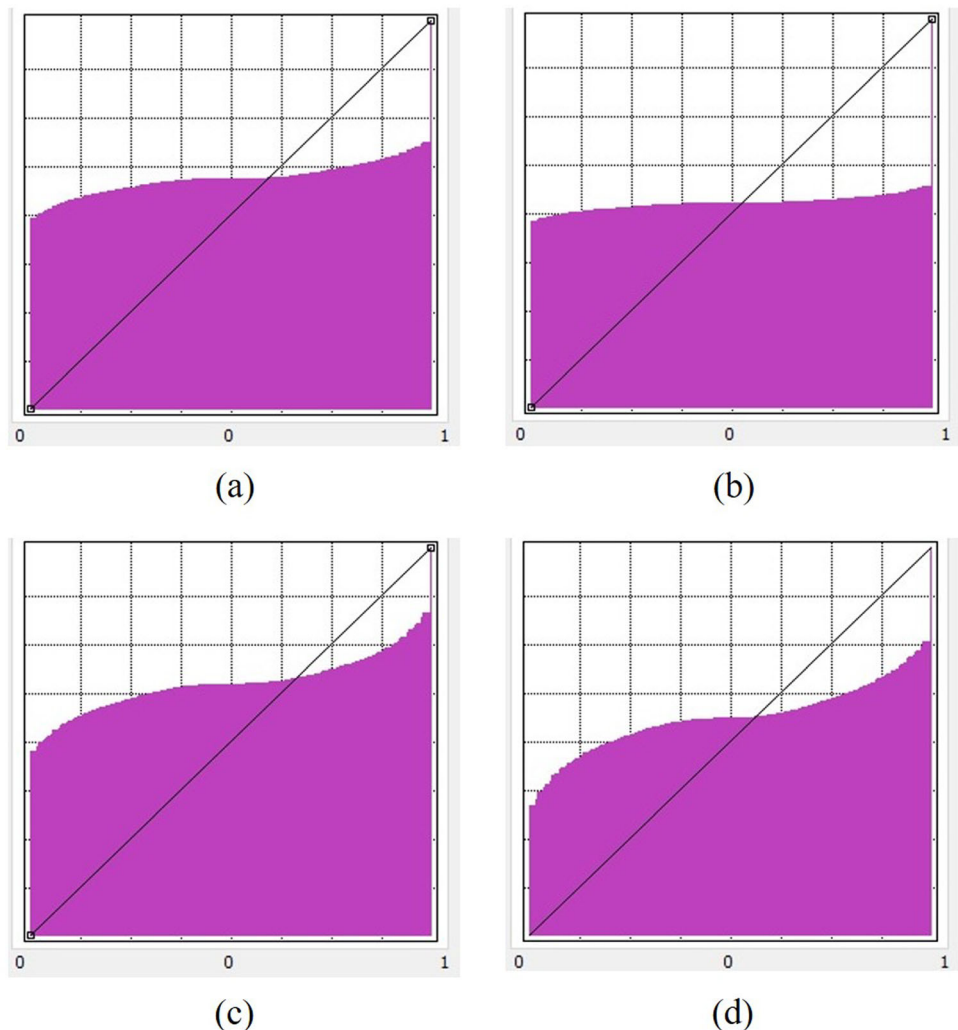
et al. 2021). As a consequence, several tactics are regularly developed globally to manage these challenges with precision and reliability (Prakash et al. 2021). To create the landslide susceptibility maps (LSMs) in this work, a hybrid model AHP-SVM is introduced by the ensemble of AHP and SVM. To pinpoint the landslide-prone areas, a variety of statistically infused, knowledge-based, probabilistic, and machine-learning methods were used. Previous studies have produced landslide susceptibility maps using a variety of quantitative and qualitative methods, including AHP (Saha et al. 2023), frequency ratio (FR) (Wubalem 2021), landslide numerical risk factor (LNRF) (Roy and Saha 2019), ANN (Selamat et al. 2022; Saha et al. 2022b), SVM (Saha et al. 2022c), logistic regression (LR) (Arabameri et al. 2019c; Chowdhuri et al. 2021b), and boosted regression tree (BRT) (Arabameri et al. 2019b) model etc. These studies identified the key landslide risk zones in the areas under consideration. Multi-hazard susceptibility modeling was adopted in Eastern Himalayas (Sikkim), where rainfall, earthquake, and high slope are the major triggering factors to happen hazards like landslides, floods, etc. (Saha et al. 2021a). Chowdhuri et al. (2021a) predict landslide susceptibility using stand-alone model LR, FR, BRT, and ensemble model BRT-FR and BRT-TR at the upper Rangit basin in Sikkim, India. This model's output was then validated with the SRC-AUC curve. ROC of stand-alone model FR, LR, and BRT are 90.1%, 86.8%, and 88.5%, respectively, and with ensemble model BRT-LR and BRT-FR get ROC of SRC-AUC are 92.4% and 94.3% respectively where it was clearly visible that ensemble methods get higher accuracy than stand-alone methods. A hybrid data mining approach was attempted by Islam et al. (2022) to generate landslide susceptibility mapping in the complex mountain region of Sikkim, India, where stand-alone method RF, Alternating Decision Tree (ADTree), and Quantum-Principle Swarm Optimization (QPSO) and ensemble approaches QPSO-ADTree were adopted, and the result shows that QPSO-ADTree outperforms others models as AUC value consists of 88.20%, 82.90%, 85.4%, and 87.60% respectively. A number of LCFs interact with triggering factors and help initiate landslides. It is crucial to choose appropriate LCFs in order to develop an accurate landslide susceptibility model. Models with high predictive potential and low error are therefore generated. There are various LCFs, and they vary based on local geographic characteristics. These landslide indicators relate to the geological aspect, climatic characteristics, and geomorphological factors that influence landslides. The area's border is dominated by a convex slope, the center is dominated by a straight slope, and a concave slope dominates the bottom part. There are no such detailed guidelines for choosing LCFs because of the various indicators and local characteristics (Park and Kim

2019). It took a lot of time and effort to choose the most appropriate LCFs. One of these projects is the multi-collinearity test, which seeks out relationships between LCFs that can impact the model's accuracy. Ten LCFs were chosen as a separate criterion in the present research to assess the study area's susceptibility to landslides. The VIF was used to assess the LCFs' multi-collinearity. Our findings show that the variables we chose don't have multicollinear behavior. As a result, the models included all of the factors.

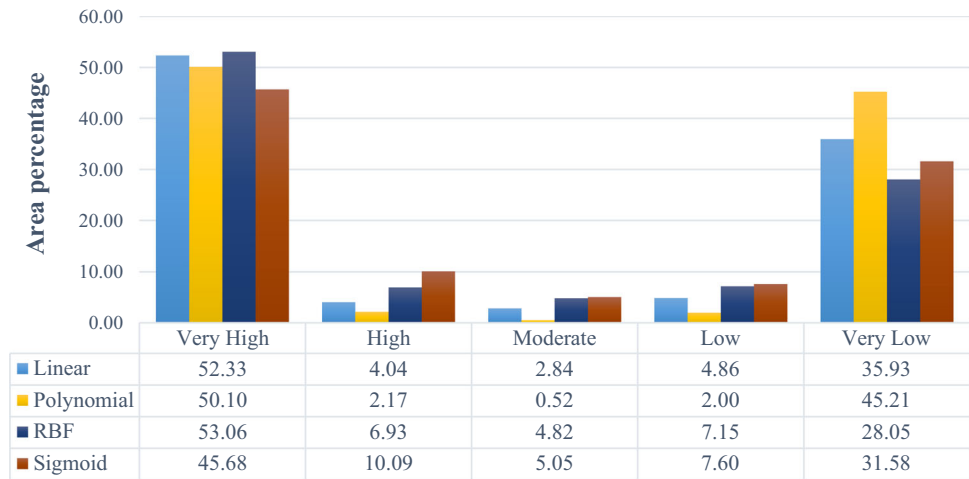
The research results show that AHP and SVM are reliable techniques for mapping landslide susceptibility zones. The AHP-SVM model uses four kernel-based algorithms: linear, polynomial, RBF, and sigmoid. The study area was divided into separate landslide susceptibility zones using ten conditioning factors, including aspect, slope, rainfall, drainage density, curvature, lineament density, soil texture, geomorphology, lithology, and LULC. The statistically generated LSMs seen in Fig. 6a–d were observed by the landslide susceptibility index (LSI) values. The mean LSI

value of AHP-SVM\_Linear was 0.41, with a standard deviation (Sd) of 0.47, while the lowest and highest LSI values were 0 and 1, respectively (Fig. 8a). AHP-SVM\_Polynomial has mean and Sd values of 0.48 and 0.49, respectively (Fig. 8b). The LSI values for the mean, s.d, min, and max of AHP-SVM\_RBF were 0.35, 0.43, 0, and 1, respectively (Fig. 8c). AHP-SVM\_Sigmoid mean and standard deviation values are 0.44 and 0.45 (Fig. 8d). According to the histogram profile, the LSI was separated into a number of zones. The histogram pattern, which offers statistical data about the cell value of LSI, illustrates the frequency of scatter data. The histogram showed that the spread values were unevenly distributed. Hence the natural break classification method was utilized for zonation mapping. Five classifications of landslide risk zones were identified and mapped as a result: very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility, and very high susceptibility (Fig. 6a–d). The analysis area percentage of the very low susceptibility zone for AHP-SVM\_Linear, AHP-SVM\_Polynomial, AHP-

**Fig. 8** Histogram analysis of different hybrid models: **a.** AHP-SVM\_Linear; **b.** AHP-SVM\_Polynomial, **c.** AHP-SVM\_RBF, **d.** AHP-SVM\_Sigmoid



**Fig. 9** Graph displaying the area % distribution for each model's classifications



SVM\_RBF, and AHP-SVM\_Sigmoid is 35.93%, 45.21%, 28.05%, and 31.58%, respectively and for very high susceptibility zone the area percentage for AHP-SVM\_Linear, AHP-SVM\_Polynomial, AHP-SVM\_RBF, and AHP-SVM\_Sigmoid is 52.33%, 50.10%, 53.06%, and 45.68%, respectively (Fig. 9). The percentage of the analysis area falls in the moderate to very high range for AHP-SVM\_Linear, AHP-SVM\_Polynomial, AHP-SVM\_RBF, and AHP-SVM\_Sigmoid is, respectively, 59.21%, 52.79%, 64.81%, and 60.82% and for AHP-SVM\_Linear, AHP-SVM\_Polynomial, AHP-SVM\_RBF, and AHP-SVM\_Sigmoid, the area falls in very low to low susceptibility zone is 40.79%, 47.21%, 35.19%, and 39.18%, respectively. As a result, we may assume that the whole study area is in a moderate zone.

The ROC-AUC was also used to evaluate the reliability of the findings using four hybrid susceptibility maps. The Geological Survey of India's public data was used to create a map of the inventory of landslides. One hundred fourteen landslides were observed in this area; 79 out of 114 (70%) were used as training data, and the rest 35 landslide locations were used as testing data. The AUC values of 86.2%, 74.2%, 73.2%, and 85.6 percent for AHP-SVM\_Sigmoid, AHP-SVM\_Linear, AHP-SVM\_Polynomial, and AHP-SVM\_RBF models, respectively, suggest that the maps produced by the Sigmoid kernel model is more accurate than the maps produced by the Linear, Polynomial, and RBF (Fig. 7). This result is helpful in an emergency since timing is a crucial component for hazard studies. It is reasonable to draw a final conclusion that the accuracy of all hybrid models is quite good. Improved accuracy, robustness, and handling of complex decision-making scenarios and intricate data relationships are the major advantages of the AHP-SVM model, as AHP is a tool for dealing with complex decision-making problems, and SVM is effective in handling nonlinear relationships between datasets. Though, AHP and SVM in an ensemble

method require careful training and optimization of the individual models as well as determining the appropriate weights or combination strategy. This process can be increased complexity as well as time-consuming and may require extensive experimentation and tuning. We recommend using AHP-SVM\_Linear, AHP-SVM\_Polynomial, AHP-SVM\_RBF, and AHP-SVM\_Sigmoid models in landslide investigations because, for risk mitigation and planning of disaster management, they can provide precise and comprehensive output. The AHP-SVM\_Sigmoid model, which is also suitable and promising, may be used to build a map of the susceptibility of landslides in a particular location.

## 5 Conclusion

In mountainous terrain, a landslide is one of the most common and worst natural hazards. Landslides threaten the severely damaged mountainous terrain due to natural (such as earthquakes and climate change) and man-made activities. These threats are turning West Bengal's northern Himalayan regions into a never-ending nightmare for the locals. Landslides are hazardous natural disasters that result in human fatalities and extensive damage to roads, homes, gardens, and agricultural land. A temporary and long-term solution to lessening the risk of landslides in this region is required to protect lives and property. In order to support future infrastructural development and urban planning, we must identify the vulnerable areas and map them. LSM may thus be an essential strategic tool for assessing risk management in hilly terrain. MCDA models have produced amazingly accurate and efficient findings for landslide susceptibility evaluations for many years. Machine learning (ML) approaches are advanced techniques developed recently. The AHP and SVM models were combined in this work to create a hybrid AHP-SVM model, which produced

LSMs for selected part of Darjeeling districts. The hybrid approach is a suitable technique for landslide susceptibility mapping that yields superior outcomes than a solo model. The four LSMs created in this research were divided into five groups according to their sensitivity to landslides: very low, low, moderate, high, and very high.

The following models represented the very high susceptibility class: 31.5% AHP-SVM\_Sigmoid, 35.93% AHP-SVM\_Linear, 45.21% AHP-SVM\_Polynomial, and 28.05% AHP-SVM\_RBF. ROC techniques were used to verify the ensemble model findings. The validation techniques validated the quality and appropriateness of the LSMs generated by the AHP-SVM\_Sigmoid, AHP-SVM\_Linear, AHP-SVM\_Polynomial, and AHP-SVM\_RBF hybrid methods. The AHP-SVM\_Sigmoid model was more accurate than other ensemble models among the four hybrid models. Landslide hazards may be decreased by locating faults and weak geological zones, managing drainage effectively, and implementing afforestation initiatives. Overall, this study performed a great job of locating the optimal planning and development zones in the mountainous Darjeeling districts of West Bengal, India, as well as the landslide-prone locations. This study's ability to forecast landslides' timing, severity, or frequency was one of its limitations. In addition, future research in the domain of LSM study should test and use deep learning models or hybrid-constructed deep learning methods by increasing adequate datasets. The newly introduced ensemble model can be used globally, where we need to segregate different classification zones from the aspect of socio-economic conditions. Darjeeling and its surroundings are one of the tourist hotspots of West Bengal, India, and tourism is one of the major economic sources for Darjeeling and its surroundings. So, all of the LSMs developed in this study may eventually be used by decision-makers, land-use planners, and government and non-government organizations as helpful tools to maximize infrastructural and socio-economic-development, various resource management, and human activity in the study area.

**Acknowledgements** The Department of Mining Engineering at the Indian Institute of Technology (Indian School of Mines), Dhanbad, provided the optimal environment for the study, for which the authors are grateful.

**Author contributions** All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by AS. The first draft of the manuscript was written by AS, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This research received no external funding.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

- Achour Y, Boumezeur A, Hadji R et al (2017) Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria. *Arab J Geosci*. <https://doi.org/10.1007/s12517-017-2980-6>
- Ahmed B (2015) Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides* 12:1077–1095. <https://doi.org/10.1007/s10346-014-0521-x>
- Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. *Bull Eng Geol Env* 58:21–44. <https://doi.org/10.1007/s100640050066>
- Arabameri A, Pradhan B, Rezaei K (2019a) Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J Environ Manag* 232:928–942. <https://doi.org/10.1016/j.jenvman.2018.11.110>
- Arabameri A, Pradhan B, Rezaei K et al (2019b) GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. *J Mt Sci* 16:595–618. <https://doi.org/10.1007/s11629-018-5168-y>
- Arabameri A, Rezaei K, Cerda A et al (2019c) GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci Total Environ* 658:160–177. <https://doi.org/10.1016/j.scitotenv.2018.12.115>
- Arabameri A, Roy J, Saha S et al (2019d) Application of probabilistic and machine learning models for groundwater potentiality mapping in Damghan sedimentary plain. *Iran Remote Sens* 11:3015
- Arabameri A, Karimi-Sangchini E, Pal SC et al (2020a) Novel credal decision tree-based ensemble approaches for predicting the landslide susceptibility. *Remote Sens* 12:3389
- Arabameri A, Pradhan B, Rezaei K et al (2020b) An ensemble model for landslide susceptibility mapping in a forested area. *Geocarto Int* 35:1680–1705. <https://doi.org/10.1080/10106049.2019.1585484>
- Azarafza M, Azarafza M, Akgün H et al (2021) Deep learning-based landslide susceptibility mapping. *Sci Rep* 11:1–16
- Basu T, Pal S (2019) RS-GIS based morphometrical and geological multi-criteria approach to the landslide susceptibility mapping in Gish River Basin, West Bengal, India. *Adv Space Res* 63:1253–1269. <https://doi.org/10.1016/j.asr.2018.10.033>
- Bera A, Mukhopadhyay BP, Das D (2019) Landslide hazard zonation mapping using multi-criteria analysis with the help of GIS techniques: a case study from Eastern Himalayas, Namchi, South Sikkim. *Nat Hazards* 96:935–959. <https://doi.org/10.1007/s11069-019-03580-w>
- Bhandari RK (2006) The Indian landslide scenario, strategic issues and action points. *India disaster management congress*, New Delhi, pp 29–30
- Carabella C, Cinosi J, Piattelli V et al (2022) Earthquake-induced landslides susceptibility evaluation: a case study from the Abruzzo region (Central Italy). *CATENA* 208:105729
- Chakraborty R et al (2022) Novel ensemble approach for landslide susceptibility index assessment in a mountainous environment of India. *Geocarto International*
- Chang K-T, Merghadi A, Yunus AP et al (2019) Evaluating scale effects of topographic variables in landslide susceptibility



- models using GIS-based machine learning techniques. *Sci Rep* 9:1–21
- Chen Y, Khan S, Paydar Z (2010) To retire or expand? A fuzzy GIS-based spatial multi-criteria evaluation framework for irrigated agriculture. *Irrig Drain* 59:174–188. <https://doi.org/10.1002/ird.470>
- Chowdhuri I, Pal SC, Arabameri A et al (2020) Ensemble approach to develop landslide susceptibility map in landslide dominated Sikkim Himalayan region, India. *Environ Earth Sci* 79:1–28
- Chowdhuri I, Pal SC, Chakraborty R et al (2021a) Torrential rainfall-induced landslide susceptibility assessment using machine learning and statistical methods of eastern Himalaya. *Nat Hazards* 107:697–722
- Chowdhuri I, Pal SC, Chakraborty R et al (2021b) Spatial prediction of landslide susceptibility using projected storm rainfall and land use in Himalayan region. *Bull Eng Geol Env* 80:5237–5258
- Chowdhuri I, Pal SC, Janizadeh S et al (2022a) Application of novel deep boosting framework-based earthquake induced landslide hazards prediction approach in Sikkim Himalaya. *Geocarto Int* 1–27(37):12509
- Chowdhuri I, Pal SC, Saha A et al (2022b) Mapping of earthquake hotspot and coldspot zones for identifying potential landslide hotspot areas in the Himalayan region. *Bull Eng Geol Env* 81:257
- Chowdhuri I, Pal SC, Saha A et al (2022c) Application of novel framework approach for assessing rainfall induced future landslide hazard to world heritage sites in Indo-Nepal-Bhutan Himalayan region. *Geocarto Int* 37:1–35
- Chowdhuri I, Pal SC, Saha A et al (2022d) Profitable agricultural land use planning in a red and lateritic soil of subtropical environment using field-based index of crop suitability (ICS). *Geocarto Int* 37:1–22
- Chowdhuri I, Pal SC, Saha A et al (2022e) Field based index of land suitability (ILS): a new approach for rainfed paddy crop production in groundwater scarce region. *Geocarto Int* 37:1–24
- Dang V-H, Hoang N-D, Nguyen L-M-D et al (2020) A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. *Forests* 11:118
- Das S, Sarkar S, Kanungo DP (2022) GIS-based landslide susceptibility zonation mapping using the analytic hierarchy process (AHP) method in parts of Kalimpong Region of Darjeeling Himalaya. *Environ Monit Assess*. <https://doi.org/10.1007/s10661-022-09851-7>
- Das S, Sarkar S, Kanungo DP (2023) A critical review on landslide susceptibility zonation: recent trends, techniques, and practices in Indian Himalaya. Springer, Netherlands
- Devkota KC, Regmi AD, Pourghasemi HR et al (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. *Nat Hazards* 65:135–165. <https://doi.org/10.1007/s11069-012-0347-6>
- Dou J, Yunus AP, Tien Bui D et al (2019) Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. *Remote Sens* 11:638
- Fan X, Scaringi G, Korup O et al (2019) Earthquake-induced chains of geologic hazards: patterns, mechanisms, and impacts. *Rev Geophys* 57:421–503
- Feizizadeh B, Blaschke T (2013) GIS-multicriteria decision analysis for landslide susceptibility mapping: comparing three methods for the Urmia lake basin, Iran. *Nat Hazards* 65:2105–2128. <https://doi.org/10.1007/s11069-012-0463-3>
- Fell R, Whitt G, Miner T, Flentje P (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102:83–84
- Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans Geosci Remote Sens* 42:1335–1343
- Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. *Nat Hazard* 18:2161–2181
- Gautam P, Kubota T, Sapkota LM, Shinohara Y (2021) Landslide susceptibility mapping with GIS in high mountain area of Nepal: a comparison of four methods. *Environ Earth Sci* 80:1–18
- Ghosh S, Carranza EJM, van Westen CJ et al (2011) Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). *Geomorphology* 131:35–56
- Ghosh S, van Westen CJ, Carranza EJM et al (2012) Generating event-based landslide maps in a data-scarce Himalayan environment for estimating temporal and magnitude probabilities. *Eng Geol* 128:49–62. <https://doi.org/10.1016/j.enggeo.2011.03.016>
- Greco R, Sorriso-Valvo M, Catalano E (2007) Logistic regression analysis in the evaluation of mass movements susceptibility: the Aspromonte case study, Calabria, Italy. *Eng Geol* 89:47–66
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216
- Guzzetti F, Mondini AC, Cardinali M et al (2012) Landslide inventory maps: New tools for an old problem. *Earth Sci Rev* 112:42–66
- Haque U, Da Silva PF, Devoli G et al (2019) The human cost of global warming: deadly landslides and their triggers (1995–2014). *Sci Total Environ* 682:673–684
- Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. *Int J Remote Sens* 23:725–749
- Islam ARMT, Saha A, Ghose B et al (2022) Landslide susceptibility modeling in a complex mountainous region of Sikkim Himalaya using new hybrid data mining approach. *Geocarto Int* 37:9021–9046
- Kalantar B, Ueda N, Saeidi V et al (2020) Landslide susceptibility mapping: machine and ensemble learning based on remote sensing big data. *Remote Sensing* 12:1737. <https://doi.org/10.3390/rs12111737>
- Kanungo DP, Arora MK, Sarkar S, Gupta RP (2006) A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng Geol* 85:347–366. <https://doi.org/10.1016/j.enggeo.2006.03.004>
- Kanwal S, Atif S, Shafiq M (2017) GIS based landslide susceptibility mapping of northern areas of Pakistan, a case study of Shigar and Shyok Basins. *Geomat Nat Haz Risk* 8:348–366. <https://doi.org/10.1080/19475705.2016.1220023>
- Kjekstad O, Highland L (2009) Economic and social impacts of landslides. In: *Landslides—disaster risk reduction*. Springer, pp 573–587
- Kolat C, Ulusay R, Suzen ML (2012) Development of geotechnical microzonation model for Yenisehir (Bursa, Turkey) located at a seismically active region. *Eng Geol* 127:36–53. <https://doi.org/10.1016/j.enggeo.2011.12.014>
- Kumar D, Thakur M, Dubey CS, Shukla DP (2017) Landslide susceptibility mapping and prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* 295:115–125
- Lee S, Sambath T (2006) Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ Geol* 50:847–855
- Li B, Wang N, Chen J (2021) GIS-based landslide susceptibility mapping using information, frequency ratio, and artificial neural

- network methods in Qinghai Province, Northwestern China. *Adv Civil Eng* 2021:1–14. <https://doi.org/10.1155/2021/4758062>
- Ma J, Liu X, Niu X et al (2020) Forecasting of landslide displacement using a probability-scheme combination ensemble prediction technique. *Int J Environ Res Public Health* 17:1–23. <https://doi.org/10.3390/ijerph17134788>
- Mallick J, Singh RK, AlAwadh MA et al (2018) GIS-based landslide susceptibility evaluation using fuzzy-AHP multi-criteria decision-making techniques in the Abha Watershed, Saudi Arabia. *Environmental Earth Sciences* 77:1–25. <https://doi.org/10.1007/s12665-018-7451-1>
- Mandal B, Mandal S (2018) Analytical hierarchy process (AHP) based landslide susceptibility mapping of Lish river basin of eastern Darjeeling Himalaya, India. *Adv Space Res* 62:3114–3132. <https://doi.org/10.1016/j.asr.2018.08.008>
- Mandal K, Saha S, Mandal S (2021) Applying deep learning and benchmark machine learning algorithms for landslide susceptibility modelling in Rorachu river basin of Sikkim Himalaya. *India Geosci Front* 12:101203
- Mandal S, Mondal S (2019) Machine learning models and spatial distribution of landslide susceptibility. In: *Geoinformatics and modelling of landslide susceptibility and risk*. Springer, pp 165–175
- Merghadi A, Yunus AP, Dou J et al (2020) Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. *Earth Sci Rev*. <https://doi.org/10.1016/j.earscirev.2020.103225>
- Mokhtari M, Abedian S (2019) Spatial prediction of landslide susceptibility in Taleghan basin, Iran. *Stoch Environ Res Risk Assess* 33:1297–1325
- Nhu V-H, Mohammadi A, Shahabi H et al (2020) Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. *Int J Environ Res Public Health* 17:4933
- Pal SC, Chowdhuri I (2019) GIS-based spatial prediction of landslide susceptibility using frequency ratio model of Lachung River basin, North Sikkim, India. *SN Appl Sci* 1:1–25
- Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Appl Sci* 9:942
- Petley D (2012) Global patterns of loss of life from landslides. *Geology* 40:927–930
- Pham BT, Jaafari A, Prakash I, Bui DT (2019) A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bull Eng Geol Env* 78:2865–2886. <https://doi.org/10.1007/s10064-018-1281-y>
- Pham BT, Prakash I, Dou J et al (2020) A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int* 35:1267–1292
- Prakash N, Manconi A, Loew S (2021) A new strategy to map landslides with a generalized convolutional neural network. *Sci Rep* 11:1–15
- Rawat A, Kumar D, Chatterjee RS, Kumar H (2022) A GIS-based liquefaction susceptibility mapping utilising the morphotectonic analysis to highlight potential hazard zones in the East Ganga plain. *Environ Earth Sci* 81:358. <https://doi.org/10.1007/s12665-022-10468-9>
- Roccati A, Paliaga G, Luino F et al (2021) GIS-based landslide susceptibility mapping for land use planning and risk assessment. *Land* 10:162
- Roy J, Saha S (2019) Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenviron Disast* 6:1–18. <https://doi.org/10.1186/s40677-019-0126-8>
- Saaty TL (2008) Decision making with the analytic hierarchy process. *Int J Serv Sci* 1:83. <https://doi.org/10.1504/IJSSCI.2008.017590>
- Saha A, Pal SC, Santosh M et al (2021a) Modelling multi-hazard threats to cultural heritage sites and environmental sustainability: the present and future scenarios. *J Clean Prod* 320:128713
- Saha S, Sarkar R, Roy J et al (2021b) Measuring landslide vulnerability status of Chukha, Bhutan using deep learning algorithms. *Sci Rep* 11:1–23
- Saha A, Pal SC, Chowdhuri I et al (2022a) Understanding the scale effects of topographical variables on landslide susceptibility mapping in Sikkim Himalaya using deep learning approaches. *Geocarto Int* 37:1–27
- Saha A, Villuri VGK, Bhardwaj A (2022b) Development and assessment of GIS-based landslide susceptibility mapping models using ANN, fuzzy-AHP, and MCDA in Darjeeling Himalayas, West Bengal, India. *Land* 11:1711. <https://doi.org/10.3390/land11101711>
- Saha S, Saha A, Hembram TK et al (2022c) Novel ensemble of deep learning neural network and support vector machine for landslide susceptibility mapping in Tehri region, Garhwal Himalaya. *Geocarto Int* 37:17018–17043. <https://doi.org/10.1080/10106049.2022.2120638>
- Saha A, Villuri VGK, Bhardwaj A, Kumar S (2023) A multi-criteria decision analysis (MCDA) approach for landslide susceptibility mapping of a part of Darjeeling District in North-East Himalaya, India. *Appl Sci* 13:5062. <https://doi.org/10.3390/app13085062>
- Saha A, Villuri VGK, Bhardwaj A (2020) Assessment and preparation of landslide susceptibility zonation map by geospatial method using remote sensing and GIS. In: *ACRS 2020 - 41st Asian conference on remote sensing*. pp 461–470
- Selamat SN, Majid NA, Taha MR, Osman A (2022) Landslide susceptibility model using Artificial Neural Network (ANN) approach in langat river Basin, Selangor. *Malaysia Land* 11:833. <https://doi.org/10.3390/land11060833>
- Sengupta A, Nath SK (2022) GIS-based landslide susceptibility mapping in eastern boundary zone of northeast india in compliance with indo-burmese subduction tectonics. In: *Geospatial technology for environmental hazards*. Springer, pp 19–37
- Shi J, Zhang J, Shen C (2020) Construct and evaluate the classification models of six types of geological hazards in Bijie city, Guizhou province, China. *Nat Hazards Earth Syst Sci Discuss* 37:1–28
- Shirzadi A, Bui DT, Pham BT et al (2017) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ Earth Sci* 76:1–18. <https://doi.org/10.1007/s12665-016-6374-y>
- Skilodimou HD, Bathrellos GD, Koskeridou E et al (2018) Physical and anthropogenic factors related to landslide activity in the Northern Peloponnese. *Greece Land* 7:85
- Sujatha ER, Sridhar V (2021) Landslide susceptibility analysis: a logistic regression model case study in Coonoor. *India Hydrology* 8:41
- Tanyaş H, van Westen CJ, Allstadt KE, Jibson RW (2019) Factors controlling landslide frequency–area distributions. *Earth Surf Proc Land* 44:900–917. <https://doi.org/10.1002/esp.4543>
- Tien Bui D, Tuan TA, Klempe H et al (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13:361–378
- Van Den Eeckhaut M, Kerle N, Poesen J, Hervás J (2012) Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. *Geomorphology* 173:30–42
- van Westen CJ, van Asch TWJ, Soeters R (2006) Landslide hazard and risk zonation - Why is it still so difficult? *Bull Eng Geol Env* 65:167–184. <https://doi.org/10.1007/s10064-005-0023-0>

- Wang J, Neskovic P, Cooper LN (2007) Selecting data for fast support vector machines training. In: Trends in neural computation, pp. 61–84
- Whiteley JS, Chambers JE, Uhlemann S et al (2019) Geophysical monitoring of moisture-induced landslides: a review. *Rev Geophys* 57:106–145
- Wind Y, Saaty TL (1980) Marketing applications of the analytic hierarchy process. *Manage Sci* 26:641–658
- Wubalem A (2021) Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia. *Geoenviron Disast* 8:1. <https://doi.org/10.1186/s40677-020-00170-y>
- Xing Y, Yue J, Guo Z et al (2021) Large-scale landslide susceptibility mapping using an integrated machine learning model: a case study in the Lvliang Mountains of China. *Front Earth Sci* 9:622
- Xu C, Dai F, Xu X, Lee YH (2012) GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* 145:70–80
- Yao X, Tham LG, Dai FC (2008) Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology* 101:572–582
- Youssef AM, Pradhan B, Dikshit A et al (2022) Landslide susceptibility mapping using CNN-1D and 2D deep learning algorithms: comparison of their performance at Asir Region, KSA. *Bull Eng Geol Environ*. <https://doi.org/10.1007/s10064-022-02657-4>
- Yu W, Liu T, Valdez R et al (2010) Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak* 10:1–7
- Yu H, Ma Y, Wang L, et al (2017) A landslide intelligent detection method based on CNN and RSG\\_R. In: 2017 IEEE international conference on mechatronics and automation (ICMA). pp 40–44
- Zhao P, Masoumi Z, Kalantari M et al (2022) A GIS-based landslide susceptibility mapping and variable importance analysis using artificial intelligent training-based methods. *Remote Sens* 14:211

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.