GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 19 • Number 3 • 2025

https://doi.org/10.7494/geom.2025.19.3.115

Saurabh Kumar Anuragi[1]

Geospatial and Optimized SVM-Based Landslide Susceptibility Zonation of South District of Sikkim, India

Abstract: Landslide identification and susceptibility maps play vital roles in supporting planners and decision-makers who manage disaster risks. By providing accurate information, these maps significantly contribute to minimizing the potential losses of life and property. To create effective landslide-susceptibility models, it is essential to incorporate a combination of terrain characteristics and meteorological factors, thus enhancing our understanding and preparedness for such events. This study presents a comparative analysis of three kernel functions (linear, polynomial, and RBF) of an support vector classifier (SVC) accompanied by a grid-search in order to determine optimal hyper-parameter settings. The primary objective of this methodological framework is to ensure accurate and reliable predictions for the generation of landslide-susceptibility maps in the South District of Sikkim, India. In this investigation, 14 conditioning factors were considered, including aspect, distance to streams, distance to roads, drainage density, elevation, lithology, land use/land cover (LULC), normalized difference vegetation index (NDVI), plan curvature, profile curvature, rainfall, slope, soil type, and earthquake susceptibility. The performances of the models were evaluated using a range of metrics, including the training score, testing score, kappa, sensitivity, specificity, accuracy, and area under the curve (AUC). Optimal hyper-parameter tuning for each SVC kernel was conducted through a grid-search approach. The results indicated that the SVC_poly and SVC_rbf models surpassed the linear model, achieving accuracy and AUC values of 0.907 and 0.908, respectively, in developing susceptibility maps. Consequently, both the SVC_poly and SVC_rbf models were identified as the most reliable and effective tools for landslide-susceptibility mapping in this study, making them optimal choices for predictive analyses in this domain.

Keywords: support vector classifier, landslide-susceptibility, kernel function, Sikkim

Received: January 5, 2025; accepted: June 2, 2025

[1] Maulana Azad National Institute of Technology, Senior Research Fellow, Department of Civil Engineering, Bhopal, India, email: saurabh.1399@gmail.com, https://orcid.org/0009-0008-7320-7527

1. Introduction

Landslides signify the downward movement of soil, rock, debris, or large land masses that are influenced by gravitational forces. These phenomena are considered among the most hazardous events, posing significant threats to human life, property, and infrastructure particularly, in mountainous regions. Various factors can trigger landslides, including heavy rainfall, seismic activity, deforestation, volcanic eruptions, and human activities such as construction or excavation. Such events disturb the soil mass, rendering it unstable and increasingly susceptible to downward movement. Landslides not only disrupt local communities but also impede economic activities particularly in vital sectors such as agriculture and tourism (which are essential to regional economy) [1–3]. Marked by its rugged topography and substantial monsoon rainfall, the Southern Sikkim area of India is particularly prone to the occurrences of landslides. Consequently, it is imperative to implement effective measures to mitigate the impacts of such disasters [4]. Conducting a landslide zonation study for the area is a crucial step toward sustainable land management, this can benefit not only Southern Sikkim but also other mountainous regions worldwide.

Landslide-susceptibility mapping plays a crucial role in understanding and mitigating the risks that are associated with landslides in a given study area. This process involves categorizing the area into distinct zones specifically, those that are labeled to have very low, low, moderate, high, and very high susceptibility levels. This classification is based on the severity or probability of landslide occurrences by taking a variety of contributing factors into account, including topographical characteristics (such as slope, elevation, and drainage patterns), geological attributes (like soil type and rock stability), and anthropogenic influences (including land-use changes and infrastructure development). Researchers in this domain have explored various techniques for assessing landslide-susceptibility. These techniques can be categorized into bivariate (frequency ratio, weight of evidence, and information value) and multivariate statistical methods (logistic regression and deterministic techniques). In bivariate statistical techniques, individual causative factors are analyzed separately in order to evaluate their contributions to landslides in a specific area. This approach aims to explore correlations or comparisons between two variables [5–7]. On the other hand, multivariate statistical techniques examine the relationships among multiple causative factors simultaneously in order to understand their relative contributions to the overall landslide-susceptibility levels [8–10]. Numerous studies have been carried out that have included various statistical methods that were used for assessing landslide-susceptibility in various geographical contexts. These include fuzzy logic, weight of evidence, the analytical hierarchy process, logistic regression, neural networks, the index of entropy, and the frequency ratio. These techniques have been applied in numerous studies across countries such as Italy, China, Iran, Bangladesh and India. The research considered different numbers of conditioning factors (geological, hydrological, and topographical factors) in order to evaluate landslide-susceptibility. The application of these statistical methods covered diverse geological settings and scales from local to regional assessments. This broad application enabled comprehensive comparisons of the effectiveness of these approaches across different landscapes, thus highlighting the versatility and applicability of these statistical methods in landslide-susceptibility assessments [11–16].

In recent years, the field of landslide-susceptibility mapping has been revolutionized by advancements in machine learning and artificial intelligence. These cutting-edge technologies have introduced novel approaches that complement and, in some cases, surpass traditional statistical methods. The integration of recent advancements in geospatial technologies such as geographic information systems (GIS) and remote sensing with machine learning algorithms has transformed the landscape of landslide-susceptibility mapping. These innovations allow researchers to generate highly accurate predictive models that analyze complex datasets and identify patterns that are indicative of landslide risk. The integration of machine learning and AI with traditional statistical methods has led to the development of hybrid approaches, thus combining the strengths of both methodologies [2, 17–22]. This synergy has resulted in more accurate and reliable landslide-susceptibility maps that are capable of capturing subtle variations in landscape characteristics and their influence on slope stability. These AI-driven techniques offer several advantages, including the ability to handle large datasets, incorporate a wide range of variables, and adapt to non-linear relationships between those factors that influence landslide occurrences. Moreover, they can often provide probabilistic outputs, which are particularly valuable for risk assessment and decision-making processes.

As the field of landslide-susceptibility mapping continues to evolve, researchers are increasingly exploring the capabilities of advanced deeplearning algorithms and extensive big data analytics [23–27]. These cutting-edge innovations are anticipated to significantly improve not only spatial resolution (thus, allowing for the finer-scale mapping of hazardous areas), but also the temporal resolutions (which are critical for understanding and predicting when landslides might occur). Enhanced predictions can lead to more timely interventions and emergency responses, thus potentially saving lives and reducing economic losses. The ultimate aim of landslide-susceptibility mapping is to facilitate proactive risk assessment and support informed mitigation planning. By accurately identifying geographic areas that are prone to landslides, various stakeholders (including urban planners, engineers, geologists, and disaster management agencies) can devise effective strategies to mitigate risks. These strategies may include land-use planning, infrastructure improvements, and community-awareness initiatives that aim to minimize damage and safeguard lives. Despite these advancements, a notable gap exists in the literature concerning the hyper-parameter tuning of these machine learning algorithms. Effective hyper-parameter tuning is essential for optimizing model performance, thus ensuring stability, and increasing the reliability of predictions. Without this fine-tuning, even sophisticated algorithms may produce results that are inconsistent or misleading, this undermines their utility in real-world applications. This study aims to develop a comprehensive landslide-susceptibility zonation model for Southern Sikkim using a support vector classifier (SVC) that is optimized through a grid-search approach. This approach is intended to refine the SVC parameters systematically, thus ensuring that the resultant model is both accurate and reliable for assessing the risks of landslides in the region. By leveraging remote sensing data and geographic information systems (GIS), various conditioning factors that influence landslide occurrences are analyzed (including aspect, distance to streams, distance to roads, drainage density, elevation, lithology, land use/land cover [LULC], normalized difference vegetation index [NDVI], plan curvature, profile curvature, rainfall, slope, soil type, and earthquake susceptibility). The optimization of the SVC model will enhance its predictive performance, thus ensuring that the resulting susceptibility maps are both accurate and reliable. The objectives of this research are threefold: first, identifying and analyzing the key factors that contribute to landslide-susceptibility in Southern Sikkim; second, optimizing hyper-parameters using a grid-search and evaluating the SVC model for three kernels (linear, poly, and RBF) for the classification of landslide-susceptibility zones; and third, validating the model's performance using established statistical methods. The outcomes of this study are expected to provide valuable insights for local authorities and stakeholders in implementing effective land-use planning and disaster risk reduction strategies.

2. Study Area

The study focuses on the South District of Sikkim, a region that is characterized by diverse topography (as is illustrated in Figure 1). This district features an extensive landscape that varies in altitude from 222 m to a towering 5,712 m. The geographical coordinates that delineate the study area span a latitude range from 27°4′N to 27°32′N and a longitude range from 88°15′E to 88°35′E. Within this region, six distinct soil types can be found, namely humid acrisols, dystric cambisols, gleysols, luvi soils, lithosols, and dystric regosols. Each of these soil types plays a crucial role in the local ecosystem and influences agricultural practices, water drainage, and vegetation growth.

The South District of Sikkim experiences significant rainfall, particularly during the monsoon months (from July through September), with an average annual precipitation of approximately 1,500 mm. This heavy rainfall contributes to the region's susceptibility to landslides, which have historically caused considerable damage to property and resulted in losses of life. The frequency and severity of these landslides have underscored the need for a comprehensive assessment of the area in order to implement effective preventive measures. By thoroughly studying the geological and climatic aspects of this district, we can develop strategies to mitigate risks and protect both human and environmental resources in this vulnerable region.

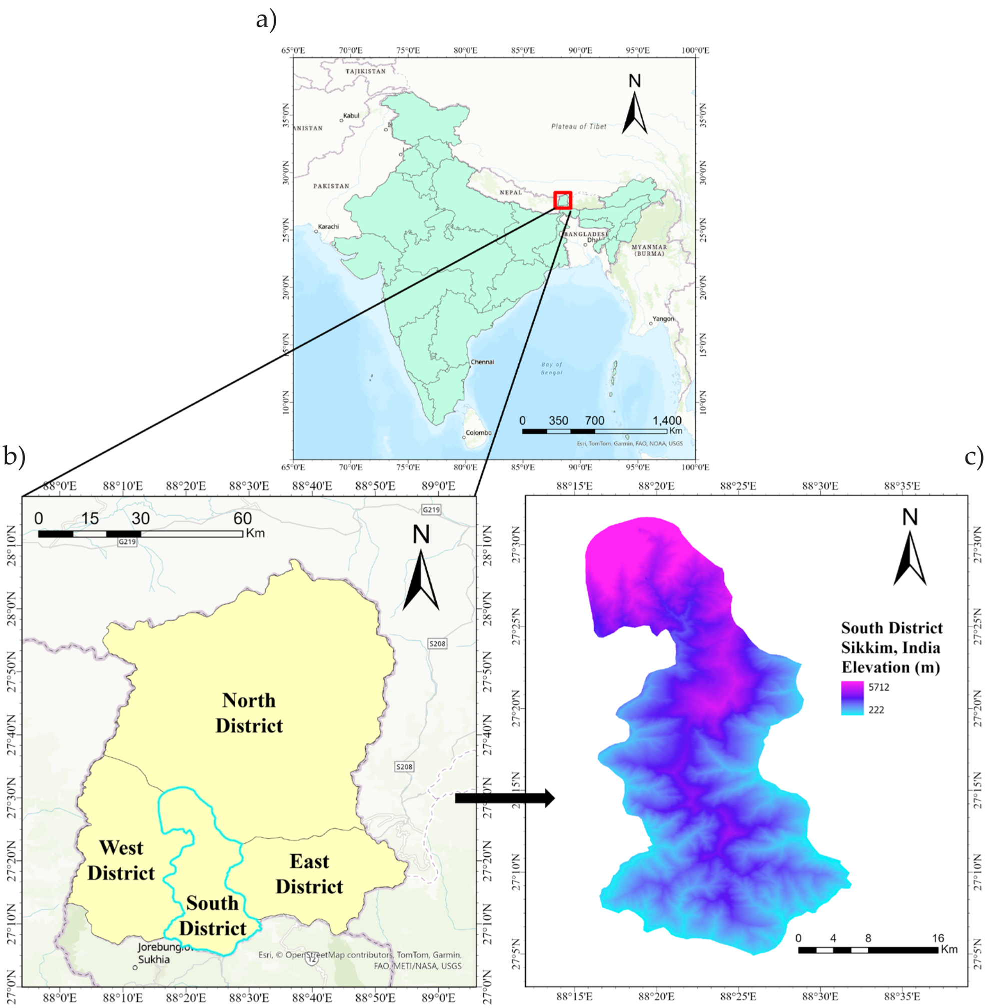


Fig. 1. Study area location map:
a) map of India; b) map of Sikkim State; c) map of South District, Sikkim, India

3. Landslide-Conditioning Factors (LCFs)

This study utilized a set of 14 conditioning factors for accurately mapping the landslide-susceptibility, including the following variables: aspect, distance to streams, distance to roads, drainage density, elevation, lithology, LULC, the NDVI, plan curvature, profile curvature, rainfall, slope, soil type, and earthquake susceptibility. These factors were meticulously analyzed in terms of their spatial distribution (as illustrated in Figure 2 on the interleaf).

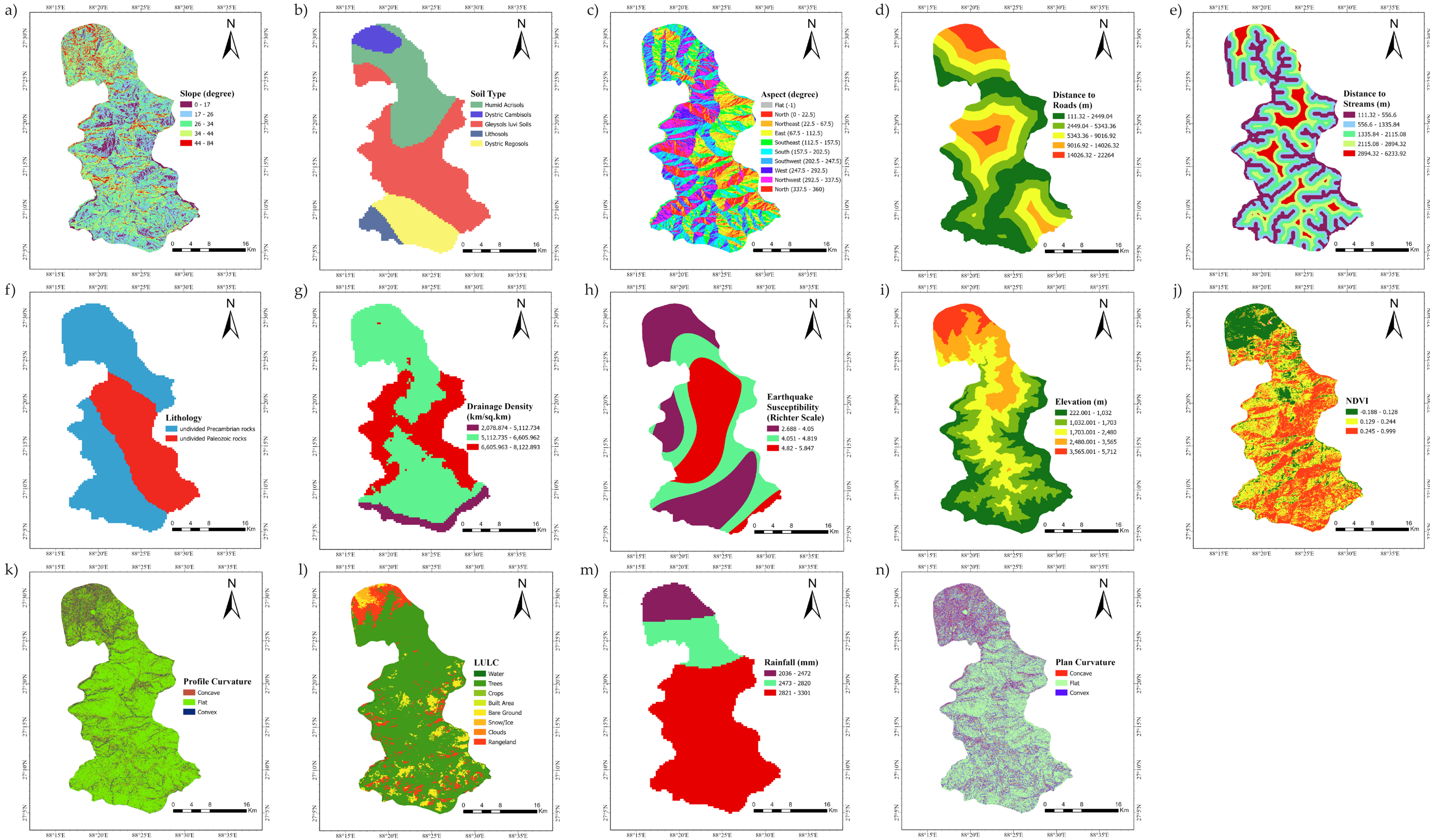


Fig. 2. Landslide-conditioning factors (LCFs) that influence landslide occurrences in study area: a) slope; b) soil type; c) aspect; d) distance to roads; e) distance to streams; f) lithology; g) drainage density; h) earthquake susceptibility; i) elevation; j) NDVI; k) profile curvature; l) LULC; m) rainfall; n) plan curvature

The aspect map (SRTM-DEM), which illustrates the orientations of the slopes, was categorized into nine distinct classifications, these include eight directional classifications: north, northeast, east, southeast, south, southwest, west, and northwest. A comprehensive understanding of slope aspect is essential, as it influences sunlight exposure, moisture retention, and soil development (each of which playing a significant role in determining vegetation cover and stability). The distance-to-streams map (https://www.hydrosheds.org/) was divided into five ranges; these indicate distances to water bodies (which have a considerable impact on soil saturation and stability). The specified ranges are as follows: (i) 111.32–556.60 m; (ii) 556.60–1335.84 m; (iii) 1335.84–2115.08 m; (iv) 2115.08–2894.32 m; and (v) 2894.32–6233.92 m. It is important to note that greater distances from streams may increase the risk of landslides due to potential soil saturation. This phenomenon occurs because river erosion removes support at the bases of slopes, while saturation weakens soil strength by increasing water content. Together, these processes greatly increase the risk and severity of landslides by making slopes more unstable and prone to failure. Similarly, the distance-to-roads map (https://diva-gis.org/) was classified into five categories based on the distances from transportation infrastructure (which frequently influences land utilization and vegetation patterns). The classifications are as follows: (i) 111.32–2449.04 m; (ii) 2449.04–5343.36 m; (iii) 5343.36–9016.92 m; (iv) 9016.92-14026.32 m; and (v) 14026.32–22264.00 m. The construction processes of roads can significantly disrupt hydrological patterns and compromise slope stability; these effects stem from both the initial construction activities and ongoing vehicular traffic. The alterations in water flow patterns, coupled with the persistent vibrations from passing vehicles, create unfavorable conditions that may increase the risk of slope instability over time. The drainage density map (SRTM-DEM) was classified into three categories: (i) 2.07–5.11 km/km^2; (ii) 5.11–6.61 km/km^2; and (iii) 6.61–8.12 km/km^2. Higher drainage density values are indicative of rapid runoff with decreased infiltration, while lower values suggest greater infiltration and slower runoff.

The elevation map (SRTM-DEM) delineated the study area into five distinct elevation ranges, which are critical for understanding the topographical factors that influence landslides: (i) 222–1032 m; (ii) 1032–1703 m; (iii) 1703–2480 m; (iv) 2480–3565 m; and (v) 3565–5712 m. Elevation is strongly correlated with climatic conditions and vegetation cover, both of these are pivotal in the dynamics of landslide occurrences. The lithology map (U.S. Geological Survey) offers a geological perspective that categorized the area into two primary types: (i) undivided Precambrian rocks; and (ii) undivided Paleozoic rocks. The nature of the underlying rock material is instrumental in determining slope stability. The LULC map (SRTM-DEM) was classified into eight categories: (i) water; (ii) trees; (iii) crops; (iv) built-up areas; (v) bare ground; (vi) snow/ice; (vii) clouds; and (viii) rangeland. This map provides valuable spatial information regarding land utilization, thus facilitating the tracking of changes in land use over time, the planning and designing of sustainable urban environments, the gaining of insights into areas that are vulnerable to natural disasters, and the identification of land that is suitable for large-scale projects. To evaluate vegetation health and its implications for landslide-susceptibility, the normalized difference vegetation index (NDVI) map (SRTM-DEM) was classified into three ranges: (i) from –0.188 to 0.128; (ii) from 0.129 to 0.244; and (iii) from 0.245 to 0.999. NDVI values reflect photosynthetic activity and vegetation health, with healthier vegetation offering enhanced soil stabilization. The plan-and-profile-curvature maps (SRTM-DEM) were categorized into three classifications (based on terrain convexity or concavity): (i) concave; (ii) flat; and (iii) convex. A land's curvature influences its water drainage and soil retention, significantly affecting the many risks that are associated with landslides. The rainfall map (https://mausam.imd.gov.in/) was segmented into three categories that measured average precipitation within the study area (a critical factor in landslide occurrences): (i) 2036–2472 mm; (ii) 2473–2820 mm; and (iii) 2821–3301 mm. Elevated precipitation levels can lead to increased soil saturation and heightened landslide-susceptibility. The slope map (SRTM-DEM) of the study area was classified into five distinct slope categories based on the degree of inclination: (i) 0–17° is categorized as a very low slope; (ii) 17–26° – as a low slope; (iii) 26–34° – as a moderate slope; (iv) 34–44° – as a high slope; and (v) 44–84° – as a very high slope. This classification is essential, as slope steepness serves as a significant indicator of landslide potential (however, the relationship is not directly proportional). On very steep slopes, the impact may be smaller, as rapid runoff reduces water infiltration into rock masses, which can slow the developments of landslides.

The soil type map categorized the study area into five distinct soil classifications: (i) humid acrisols; (ii) dystric cambisols; (iii) gleysols luvi; (iv) lithosols; and (v) dystric regosols. Each of these soil types possesses unique characteristics that influence moisture retention and erosion potential (which are both essential for understanding landslide behavior). Additionally, the earthquake map segmented seismic events into three magnitude ranges: (i) 2.688–4.05; (ii) 4.05–4.819; and (iii) 4.82–5.847. This recognizes the influence of seismic activity on the occurrences of landslides. The thematic maps that were developed in this study were crafted with meticulous attention to detail, thus facilitating a comprehensive analysis of landslide-susceptibility through their integration with landslide occurrence data. This holistic approach enabled a more accurate and thorough evaluation of the various factors that contribute to landslide risks within the study area.

4. Multicollinearity Analysis of Landslide Conditioning Factors

Multicollinearity analysis is a crucial step in evaluating the relationships among independent variables (often referred to as LCFs in this study). Multicollinearity occurs when strong correlations exist between LCFs, which can introduce inaccuracies and lead to instability in a predictive model. To assess the degrees of correlation among the LCFs within the study area, this investigation employed two widely recognized statistical techniques: Pearson's correlation coefficient and the variance inflation factor (VIF). Pearson's correlation coefficient measures the linear relationship between pairs of variables, while VIF quantifies how much the variance of an estimated regression coefficient increases due to multicollinearity. For this analysis, the LCFs were selected based on a Pearson's correlation threshold of 0.7 (indicating a strong linear relationship) and a maximum acceptable VIF value of 10 (suggesting that multicollinearity does not excessively influence the model).

Figures 3 and 4 present the findings for Pearson's correlation and VIF, respectively. Notably, the highest Pearson's correlation coefficient was found to be 0.49, which indicated a moderate positive correlation between elevation and the distance to streams. This correlation suggested that, as the elevation increased, the distances to streams also tended to increase (although the correlation was not strong enough to raise concerns of multicollinearity). Additionally, the maximum VIF that was calculated for one of the LCFs was 3.52 (specifically, for the elevation factor). This value fell well below the threshold of 10, further supporting the absence of problematic multicollinearity within the selected variables.

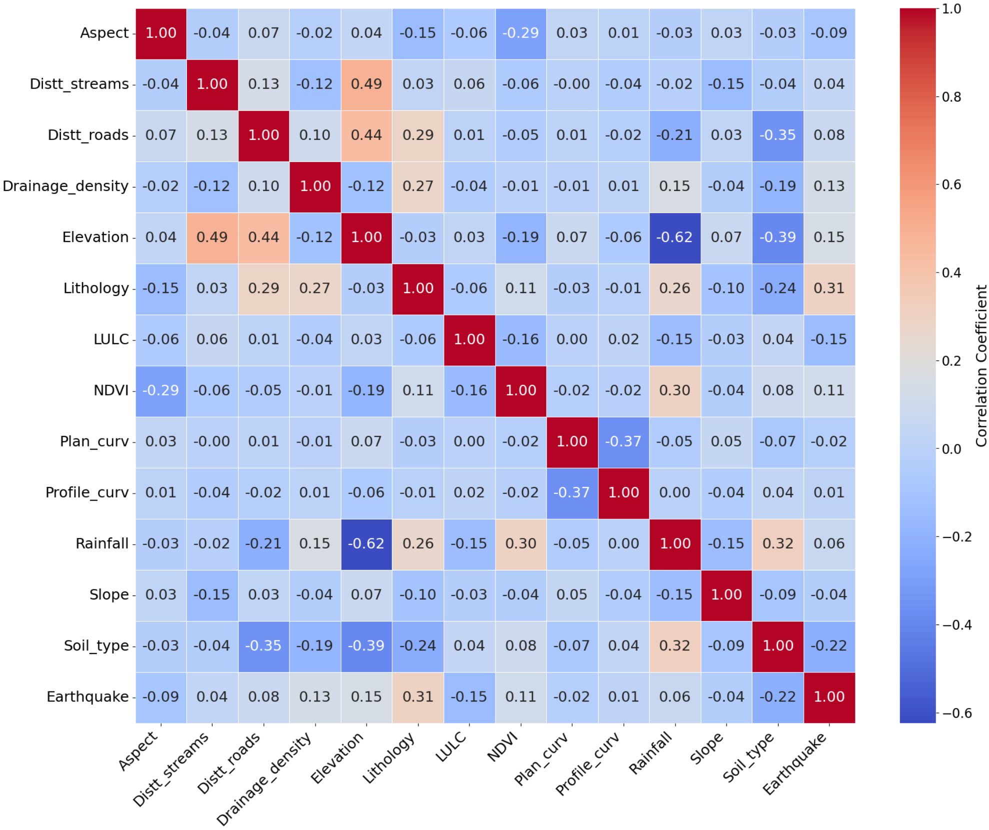


Fig. 3. Pearson correlation matrix

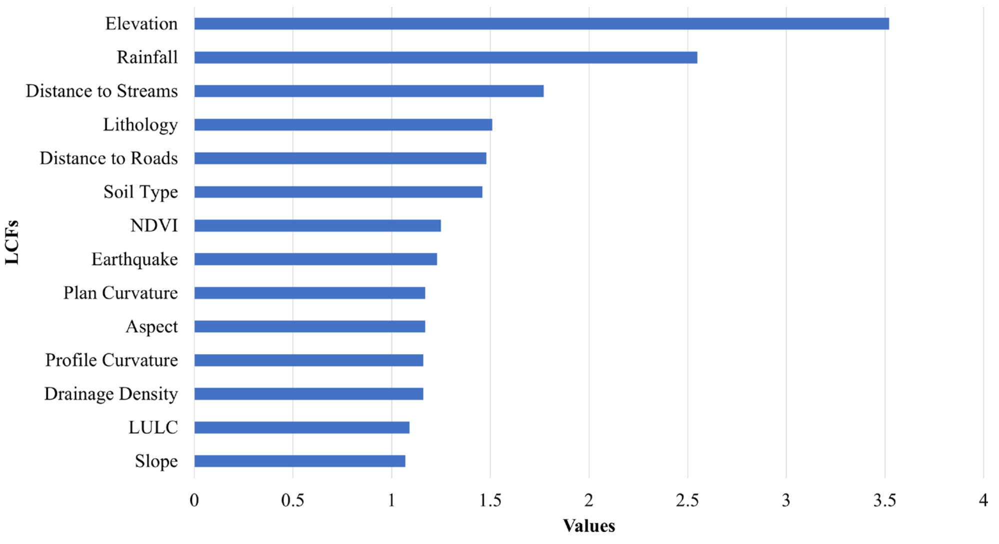


Fig. 4. VIF analyses of LCFs

The outcomes of this comprehensive multicollinearity analysis indicated that all of the proposed landslide-conditioning factors were viable for use when modeling the landslide-susceptibility within the southern region of Sikkim. The results not only affirmed the suitability of the LCFs but also enhanced the reliability of the modeling process (which aimed to understand and predict landslide occurrences in this area).

5. Landslide Inventory

The landslide data that was utilized in this analysis was obtained from the Bhukosh Geological Survey of India. The landslide inventory was comprised only of reported events (which are predominantly documented in accessible or populated areas). Consequently, it may not have comprehensively captured landslide occurrences in remote or inaccessible regions, thus leading to potential spatial bias in the dataset. This dataset encompassed a total of 196 distinct landslide data points (as illustrated in Figure 5). These data points were subsequently imported into ArcGIS (a sophisticated geographic information system) to generate polygons that served as the basis for the comprehensive dataset that was required for in-depth analysis. In order to create a balanced and robust dataset, an additional 200 non-landslide data points were randomly generated within ArcGIS. Corresponding polygons for these non-landslide points were also developed, thus ensuring that the analysis accounted for both landslide and non-landslide occurrences. By integrating the LCFs with the landslide and non-landslide data, a consolidated dataset that consisted of 2,648 data points was established for a thorough evaluation. This consolidated dataset was subsequently divided using a 70:30 ratio (with 70% allocated for training purposes, and 30% reserved for testing). This strategic division was designed to facilitate a comprehensive analysis, thus allowing for the development of predictive models while ensuring the reliability and validity of the results.

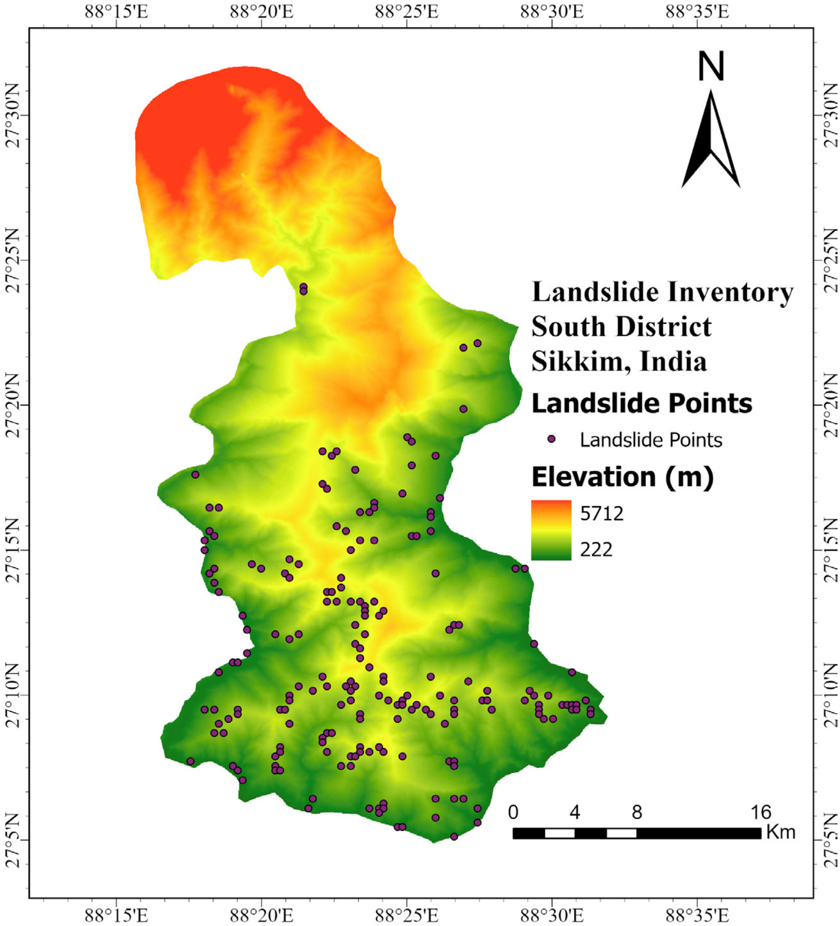


Fig. 5. Landslide point location map of South District, Sikkim, India

6. Support Vector Classifier (SVC)

Support vector classifiers are supervised learning models machine-learning algorithms that were presented by Cortes and Vapnik [28] that were based on the statistical learning theory and used to perform regression and classification analyses. The objective of SVC algorithms is to find the largest margins between two classes by hyperplanes. SVCs utilize the concept of mapping data into high-dimensional spaces (where linear classifications are carried out). SVC algorithms were developed from the optimal problem of a classification hyper-plane under linearly separable conditions. The concept of the algorithms is to maximize the intervals of training sets and minimize bounds on the generalization errors of models rather than minimizing only the mean square errors over the datasets. The kernel functions (polynomial, linear, radial basis function) make SVCs more flexible and able to handle non-linear problems. Therefore, SVCs are also used for classifying non-linear problems.

The introduction of kernel functions by Boser et al. [29] extended SVC's capability of handling non-linearly separable data. Cortes and Vapnik [28] presented the SVC formulation, showcasing its ability to find optimal separating hyperplanes with maximum margins by introducing two terms; i.e., slack variable ξ, and penalty factor C. Slack variable ξ measures the standard deviation of a data pattern from the ideal condition, whereas penalty factor C defines the trade-off having a wide margin and fewer classification errors in training data.

Key functions and conditions in SVM are as follows:

- Classification constraints:
  - for linearly separable data:

$$y_i[(w^T x_i) + b] - 1 \geq 0 \tag{1}$$

  - for linearly inseparable data:

$$y_i[(w^T x_i) + b] \geq 1 - \xi_i \tag{2}$$

- Objective function (to minimize):

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \tag{3}$$

- SVM kernel functions:
  - linear kernel:

$$k(x_i x_j) = x_i^T x_j \tag{4}$$

  - polynomial kernel:

$$k(x_i x_j) = (\Upsilon x_i^T + r)^d, \ \Upsilon > 0 \tag{5}$$

  - radial basis function (RBF) kernel:

$$k(x_i x_j) = e^{\left(\Upsilon \left\| x_i - x_j \right\|^2\right)}, \ \Upsilon > 0 \tag{6}$$

where w is an adaptive weight factor, x is an input vector, b is bias, $w^T x$ is an inner product of w and x, and Υ, r, and d are the kernel parameters.

The current study utilizes grid-search to optimize hyper-parameters of the SVC model, including *C*, gamma, kernel, and degree. This method systematically explores a range of values to find the best combination for model performance. Grid-search evaluates each parameter combination using cross-validation to ensure good generalization to unseen data, resulting in an accurate and reliable SVC model (as shown in Table 1).

Table 1. SVC hyperparameter tuning using grid search

Model	Hyperparameter range	Optimal values
SVC_linear	C = [1–200]; step size = 1	97
SVC_poly	C = [1–200]; step size = 1	34
	Degree = [1, 2, 3]	2
SVC_rbf	C = [1–200]; step size = 1	122
	Gamma = [0.001, 0.01, 0.1]	0.01

7. Methodology

The methodology that was employed in this study (depicted in Figure 6) delineates a thorough and systematic framework that was aimed at achieving the accurate and dependable mapping of landslide-susceptibility. A critical component of this process was the optimization of the hyper-parameters for the machine learning models that were used specifically for the different kernel functions of the SVC. Prior to conducting the primary analysis, it was imperative to refine these hyper-parameters in order to enhance the model's predictive accuracy. To facilitate effective hyper-parameter optimization, the study utilized a structured approach known as grid search; this method is characterized by its systematic and comprehensive nature that involves the definition of a clearly structured range of potential hyper-parameter values. The grid search process systematically evaluates each combination of these values to determine the configuration that provides the optimal performance for the specific objectives of landslide-susceptibility mapping. The study thoroughly investigated various combinations of hyper-parameters in order to ascertain the optimal settings for each machine learning model, thus ensuring their efficacy in modeling landslide-susceptibility. The specific ranges of the hyper-parameters that were considered during this tuning process are detailed in Table 1 (which contributes to the transparency and reproducibility of the methodology). Through this rigorous optimization procedure, the models were refined to achieve their peak performance. This optimization step was vital for enhancing the accuracy and reliability of the landslide-susceptibility mapping, thus establishing a solid foundation for subsequent analyses and interpretations.

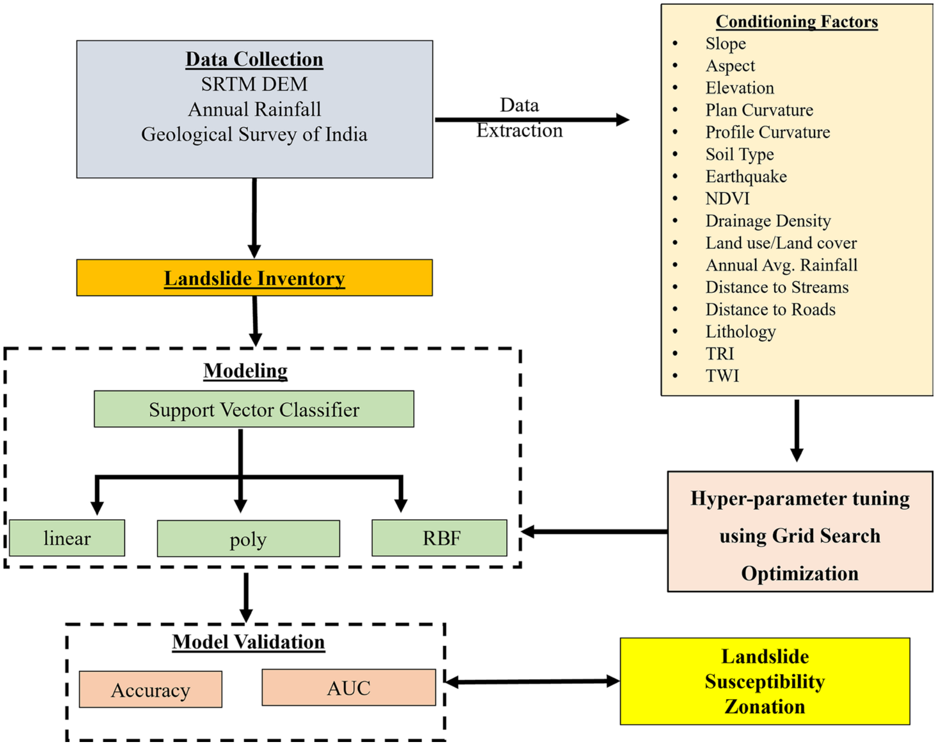


Fig. 6. Methodology of study

The weights that were assigned to each thematic layer were analyzed for the various kernel functions of the SVC (specifically, linear, polynomial, and RBF kernels). Each thematic layer received a weight on a scale from 0 to 1, with higher values indicating greater influences on the occurrences of landslides. Furthermore, the causative factors were categorized according to the severity of their roles in triggering landslides for each SVC kernel function. All of the thematic maps that depicted these causative factors were compiled in the raster format, with a pixel size of 30 m × 30 m. The final landslide-hazard-zonation map was developed through a spatial overlay analysis of each thematic layer. The comprehensive methodology that was employed in this study enhances our understanding of the factors that contribute to landslide occurrences and highlights the significance of thorough preparation in predictive modeling.

8. Results and Discussion

8.1. Model Performance Comparison

This study employed the support vector classification (SVC) machine learning technique that utilized various kernel functions (including linear, polynomial, and radial basis functions [RBF]) to develop accurate and reliable models for generating landslide-susceptibility maps. Each model underwent a comprehensive hyper-parameter optimization process that was conducted through grid-search optimization. The optimal hyper-parameters that were obtained during this process are presented in Table 1. The results were systematically compared, thus facilitating in-depth analysis of the performances of the machine learning models. These results were visually represented through corresponding susceptibility maps, thus offering valuable insights into understanding and predicting landslide occurrences. A critical component of the analyses involved evaluations of the classifier's performance that were emphasized by area under the curve (AUC) and accuracy metrics. The AUC highlighted the necessity of thorough evaluations of the accuracy of the models in the context of landslide-susceptibility assessment, thus ensuring that the findings were both reliable and actionable for the relevant stakeholders. The accuracy metric assessed the proportion of the correct predictions relative to the total number of predictions, thus providing a more dependable estimate of the model's performances by incorporating all of the data points for both training and testing. To evaluate the accuracy, the five-fold cross-validation technique was utilized, wherein the model was trained on four folds (with the remaining one designated for the testing). This procedure was repeated five times each time, with a different fold serving as the test set.

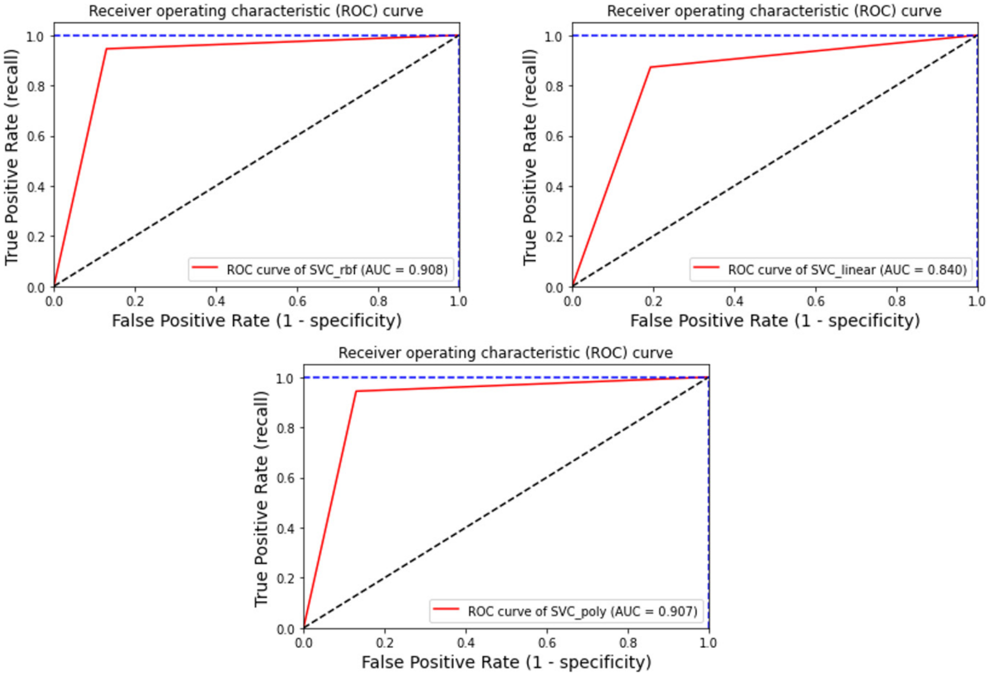


Fig. 7. ROC curves

Figure 7 illustrates the ROC curves for the various SVC kernels, thus highlighting their effectiveness in distinguishing among the classes. The AUC values, which assessed the model's discriminative capabilities, indicated that SVC_rbf and SVC_poly achieved AUC values of 0.908 and 0.907, respectively. These figures suggested a marked improvement in performance as compared to SVC_linear, which registered an AUC value of 0.84. The differences in these values reflected the inherent characteristics of the respective algorithms, including their complexity and their aptitude for capturing feature-target relationships. The elevated AUC values for SVC_rbf and SVC_poly implied superior discriminatory ability and overall performance. In addition, Figure 8 presents a broader evaluation of several metrics, including the training and testing scores, specificity, sensitivity, accuracy, and kappa coefficient, thus facilitating a comprehensive performance comparison of the models.

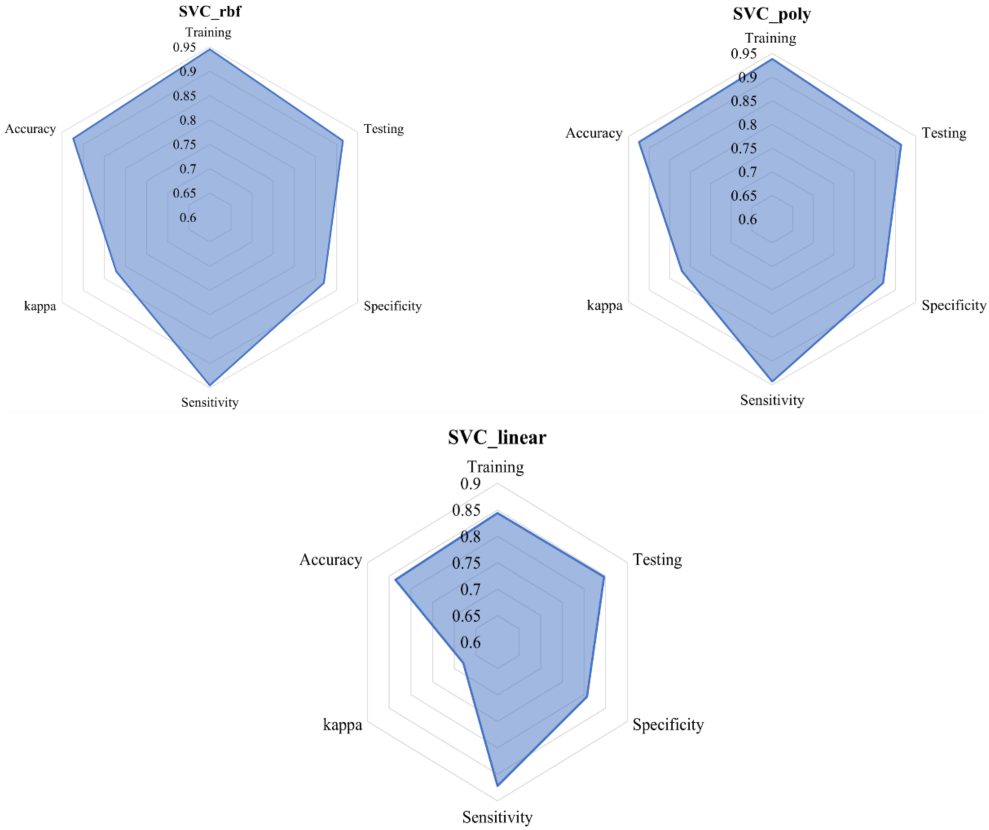


Fig. 8. Spider plots of evaluation metrics

For SVC_linear, the metrics were as follows: training score – 0.844; testing score – 0.846; specificity – 0.807; sensitivity – 0.872; accuracy – 0.836; and kappa – 0.679. In contrast, SVC_poly exhibited more favorable outcomes, with a training score – 0.938; testing score – 0.914; specificity – 0.87; sensitivity – 0.943; accuracy – 0.925; and kappa – 0.82. Likewise, SVC_rbf yielded a training score – 0.945; testing score – 0.915; specificity – 0.87; sensitivity – 0.945; accuracy – 0.924; and kappa – 0.822. These results underscored the robust performance of SVC_rbf and SVC_poly across all of the assessed metrics, thus demonstrating their capability to effectively generalize and capture any complex relationships within the data. While SVC_linear delivered adequate results, its simpler algorithm conversely evidenced a comparatively lower performance due to its limitations in modeling the non-linear relationships. Overall, this analysis highlighted the effectiveness of SVC_rbf and SVC_poly for tasks that demand precise classification and strong generalization.

This study provided a thorough evaluation of ROC curves alongside their relevant metrics in order to assess the predictive accuracies of the various models. A key component of this analysis was an examination of the confusion matrices (as are depicted in Figure 9). These matrices offered a detailed view of the classification performance by outlining the distribution of true positives, true negatives, false positives, and false negatives.

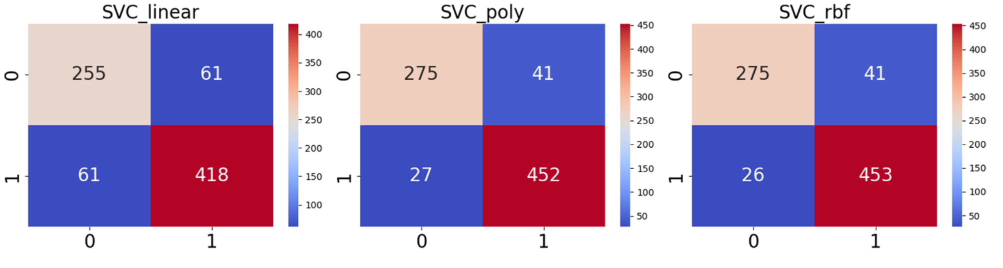


Fig. 9. Confusion matrix

The total number of misclassifications for each model served as an essential measure that reinforced the findings as related to the AUC and the overall effectiveness of the evaluated models. Among the models that were studied, SVC_linear demonstrated the highest number of misclassifications (amounting to 122); this indicated its relatively lower predictive accuracy. In contrast, SVC_poly and SVC_rbf recorded the fewest misclassifications (with totals of 68 and 67, respectively). The similarity in the misclassification rates for these two models (SVC_poly and SVC_rbf) corresponded with their higher AUC values, further validating their enhanced capabilities of distinguishing between susceptible and non-susceptible areas for landslide occurrences. This analysis not only highlighted the superior discriminatory power of SVC_poly and SVC_rbf but also underscored the practical implications of these models concerning their classification reliability. By integrating the findings from the confusion matrices with the other performance metrics, this study emphasized the robustness of these kernels in landslide-susceptibility modeling, thereby establishing their appropriateness for tasks that demand high precision and minimal error rates.

8.2. Landslide Susceptibility Zonation

The landslide-susceptibility maps that were generated using the various SVC kernels are illustrated in Figure 10. These maps provide a visual representation of how each kernel classified the study area into distinct levels of landslide-susceptibility. They offer critical insights into the spatial distribution of landslide risks, thus allowing stakeholders to identify high-risk zones and prioritize areas for targeted mitigation efforts.

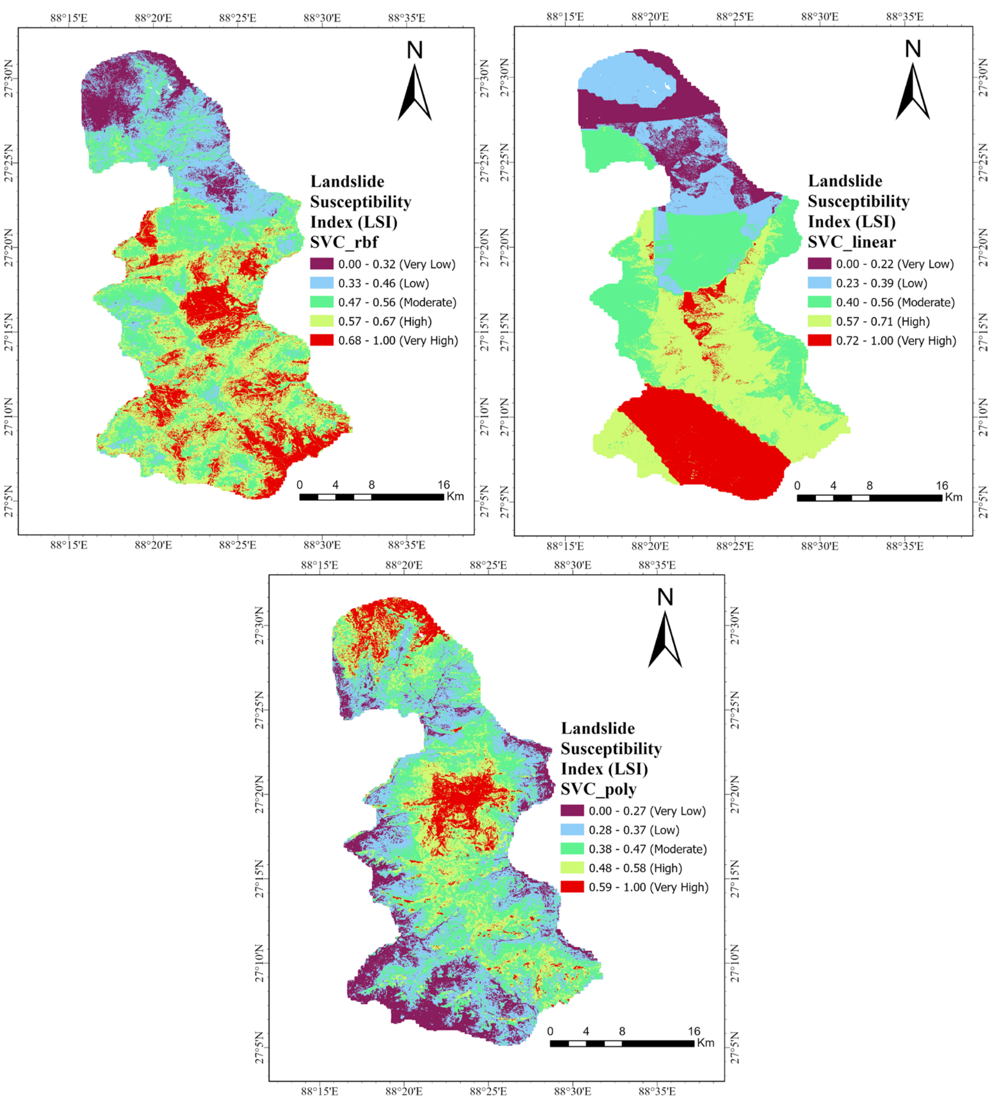
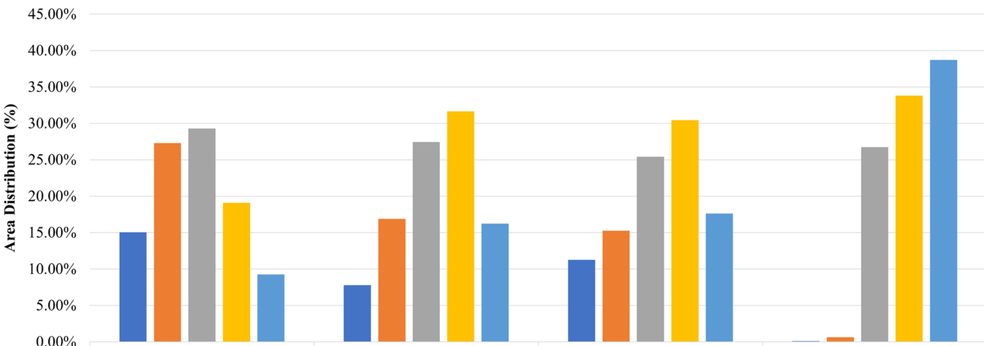


Fig. 10. Generated landslide-susceptibility maps by SVC kernels

To complement the visual analysis, Figure 11 presents the percentage-area distribution for each susceptibility classification (very low, low, moderate, high, and very high) across the different models. This quantitative assessment highlighted the degree to which each kernel predicted the susceptibility levels within the study area. For the SVC_linear kernel, the area that was classified as having very low susceptibility was the smallest (11.25% of the total area), while the high susceptibility class covered the largest proportion (30.46%). This distribution indicated that the SVC_linear kernel took a more conservative approach in designating areas as having very low risk, while a substantial portion was assigned to highersusceptibility levels. When compared to the historical landslide distribution, however (where 38.71% of the area fell under the very high class), the model still underrepresented the most critical zones.



	SVC_poly	SVC_rbf	SVC_linear	Landslide (%)
Very Low	15.04%	7.79%	11.25%	0.11%
Low	27.30%	16.87%	15.26%	0.64%
Moderate	29.30%	27.45%	25.42%	26.74%
High	19.09%	31.67%	30.46%	33.81%
Very High	9.26%	16.22%	17.62%	38.71%

Fig. 11. Area distribution of study area

The SVC_poly kernel designated the very high susceptibility class as the smallest area (9.26%), with the moderate susceptibility class being the largest (29.30%). This distribution reflected a more balanced predictive capability, thus maintaining a relatively even allocation of those areas across the moderate and high-risk categories while representing fewer extremes in the very high and very low susceptibility areas. This broader spread may have reduced false positives, but it might have also underrepresented critical areas, as only a small fraction aligned with historically high landslide densities.

For the SVC_rbf kernel, the very low susceptibility class constituted the smallest proportion (7.79%), with the high susceptibility class displaying the largest coverage (31.67%). This trend suggested a stronger correlation with the actual landslide distribution (which showed 33.81% and 38.71% in the high and very high categories, respectively), thereby supporting its observed superior performance metrics (like its higher AUC and fewer misclassifications).

As depicted in Figure 12, the final susceptibility map was derived as the average landslide-susceptibility index (LSI) of three maps (as is shown in Figure 10). The probability of landslide occurrences was quantified using the LSI, which ranged from 0 to 1. Here, a value of 0 indicated a very low probability of a landslide occurrence, while a value of 1 signified a very high probability.

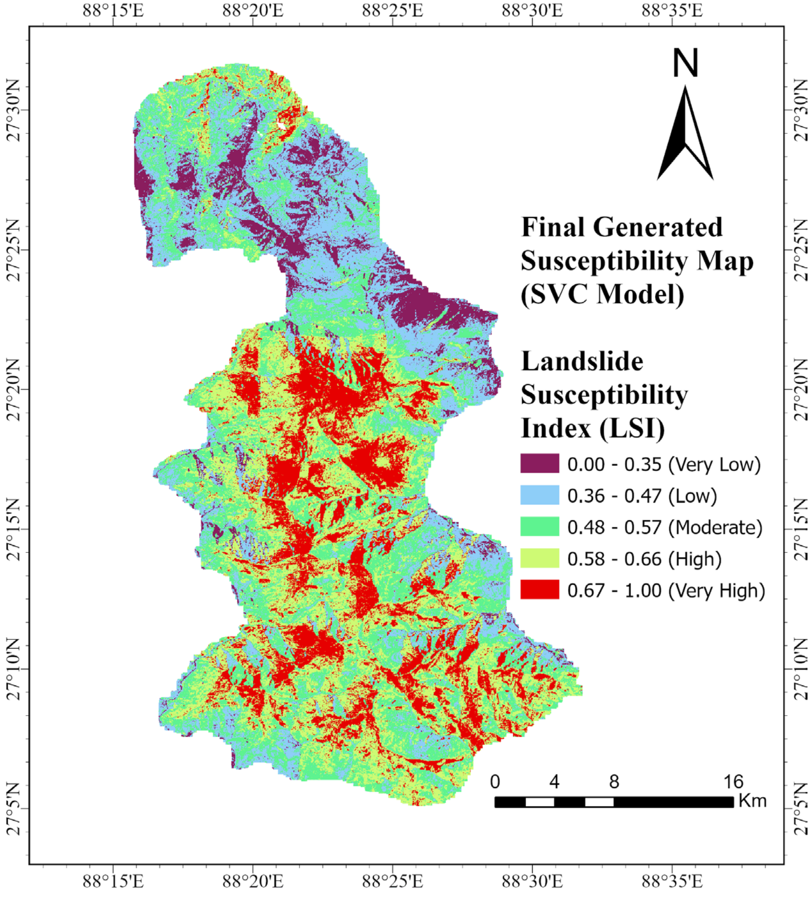


Fig. 12. Final susceptibility map of study area

The area distribution of the final susceptibility map (as shown in Figure 13) presents the percentage area distribution for each susceptibility classification. It can be observed that the area that was classified as very low susceptibility was the smallest (7.83%), while the moderate susceptibility class encompassed the largest proportion (27.31%). This distribution suggested that the final susceptibility map adopted a more conservative approach when designating areas as very low risk, while a significant portion was assigned to moderate susceptibility levels.

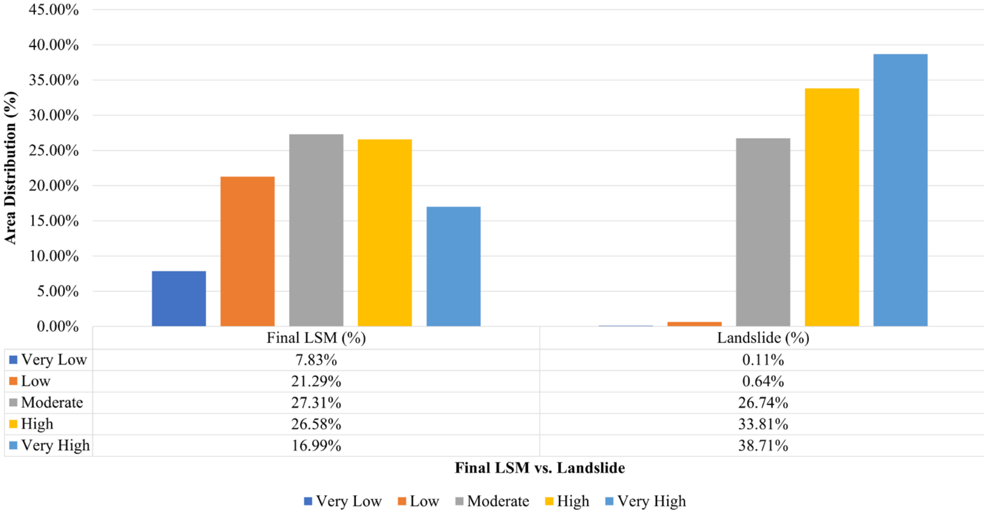


Fig. 13. Area distribution of final susceptibility map of study area

These comparative analyses of the susceptibility maps and area distributions underscored the unique predictive characteristics of each SVC kernel. By integrating both visual and quantitative insights, this study provides a comprehensive framework for evaluating landslide risks, thereby offering valuable guidance for decision-making in hazard management and risk mitigation planning.

8.3. Feature Importance Analysis

Feature importance plays a pivotal role in landslide prediction, as it involves assessing the relevance and contribution of each input feature in the decision-making process of a model. By quantifying the influence of specific features on the model's predictions, feature importance provides valuable insights into the factors that most significantly impact landslide-susceptibility. In this study, 14 features were utilized for landslide-susceptibility mapping: aspect, distance to streams, distance to roads, drainage density, elevation, lithology, LULC, the NDVI, plan curvature, profile curvature, rainfall, slope, soil type, and earthquake susceptibility. Figure 14 presents the feature importance values that were assigned by each kernel of the SVC, which were scaled between 0 and 1. These values allowed for a comparative analysis of how each kernel prioritized the input features. To facilitate the interpretation, the features were categorized based on their assigned weightages into three levels: low (0.0–0.3); moderate (0.3–0.6); and high (0.6–1.0) (as summarized in Table 2).

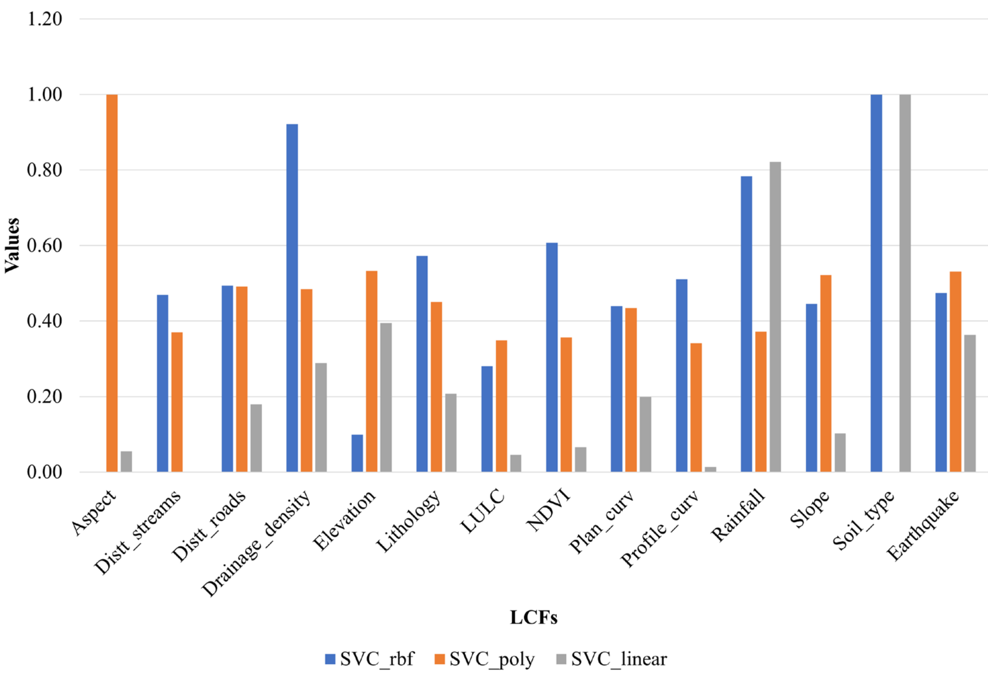


Fig. 14. Feature Importance by SVC kernels

The categorization highlighted the diversity in the feature importance across the SVC kernels, thus reflecting differences in their learning algorithms and decision-making processes.

Table 2. Feature categorization

Model	Low (0.0–0.3)	Moderate (0.3–0.6)	High (0.6–1.0)
SVC_rbf	Aspect, Distance to streams, Drainage density	Elevation, Distance to roads, NDVI	LULC, Lithology, Rainfall, Plan curvature, Soil type, Profile curvature, Slope, Earthquake susceptibility
SVC_poly	LULC, Distance to streams, Aspect	NDVI, Distance to roads, Soil type, Drainage density	Elevation, Lithology, Plan curvature, Profile curvature, Rainfall, Slope, Earthquake susceptibility
SVC_linear	Aspect, Elevation, Rainfall	Distance to streams, Earthquake susceptibility, Soil type	Distance to roads, Drainage density, Lithology, LULC, NDVI, Plan curvature, Profile curvature, Slope

The analysis of the SVC kernels revealed distinct patterns in their feature importance concerning landslide-susceptibility. In the SVC_linear kernel, features that were classified under the low-importance category included aspect, distance to streams, distance to roads, drainage density, lithology, LULC, NDVI, plan curvature, profile curvature, and slope. Elevation and earthquake susceptibility were categorized as moderate, whereas rainfall and soil type were identified as high-importance features, thus clearly reflecting their significant impacts on landslide occurrences.

Conversely, the SVC_poly kernel designated LULC, NDVI, and soil type as low-importance features. A broader range of features (including distance to streams, distance to roads, drainage density, elevation, lithology, plan curvature, profile curvature, rainfall, slope, and earthquake susceptibility) were classified as moderate. Notably, aspect was classified as high importance, thus suggesting its critical role in landslide-susceptibility mapping.

In the SVC_rbf kernel, aspect, elevation, and LULC were categorized as low-importance features. Moderate features included distance to streams, distance to roads, lithology, plan curvature, profile curvature, slope, and earthquake susceptibility, while drainage density, NDVI, rainfall, and soil type were classified as high importance. This underscored the significant influence of hydrological and vegetative factors along with soil type in determining landslide-susceptibility.

The differing feature importance across the SVC kernels highlighted the unique methodologies that were employed by each model in processing the input data. An understanding of these patterns provides valuable insights into the driving factors behind landslide-susceptibility, thus contributing to the enhancement of predictive models.

9. Conclusion

The South District of Sikkim is prone to frequent landslides; these are primarily attributed to seismic activity and rainfall. It is essential to identify the most vulnerable regions to facilitate informed decision-making and effective planning in land-use management and hazard mitigation.

This study focused on the landslide zonation of the South District of Sikkim by employing various kernel functions (linear, polynomial, and radial basis function) of a support vector classifier (SVC), utilizing a geographic information system (GIS) and remote sensing techniques. The analysis incorporated 14 conditioning factors, including aspect, distance to streams, distance to roads, drainage density, elevation, lithology, LULC, NDVI, plan curvature, profile curvature, rainfall, slope, soil type, and earthquake susceptibility. These factors were carefully selected to provide a comprehensive understanding of the region’s susceptibility to landslides.

The resulting landslide zonation categorized the study area into five classes: very low, low, moderate, high, and very high susceptibility.

The conclusions that were drawn from this study, based on the data analysis, can be summarized as follows:

- The application of grid-search for hyperparameter tuning proved to be an effective approach in optimizing the performances of SVC models. By systematically evaluating a wide range of parameter combinations, we identified optimal hyperparameter values that significantly enhanced model reliability and accuracy. The results that are presented in Table 1 illustrate how tailored adjustments to parameters such as the regularization constant (*C*), degree, and gamma could lead to specific improvements for different SVC kernels. This meticulous tuning process underscored the importance of careful parameter selection in building robust machine-learning models that are capable of generalizing well to unseen data. Future work can build on these findings by exploring additional hyperparameter ranges and more sophisticated optimization techniques to further elevate model performance.
- The performances of various models were evaluated through a comprehensive set of metrics, including confusion matrices, training and testing scores, kappa, sensitivity, specificity, accuracy, and AUC. This multifaceted approach ensured a thorough and nuanced assessment of each model's capability to accurately classify landslide-susceptibility. Among the models that were assessed, SVC_poly and SVC_rbf demonstrated superior performance when compared to SVC_linear, achieving AUC values of 0.907 and 0.908, respectively, along with respective accuracy rates of 0.925 and 0.924. Additionally, the numbers of misclassifications for SVC_poly (68) and SVC_rbf (67) were significantly lower than those of SVC_linear (122), thereby reinforcing the reliability and precision of these models in identifying areas that are prone to landslides. The findings of this study highlighted the effectiveness of machine-learning techniques (particularly, SVC_poly and SVC_rbf) in producing accurate and dependable landslide-susceptibility maps. The final susceptibility map was produced by averaging the LSIs of the generated maps using the three kernels of the SVC. This map (Fig. 12) illustrated the overall susceptibility of the region to landslides based on a comprehensive analysis of multiple factors. Those areas with higher susceptibility were distinctly marked, thus offering critical insights into land-use planning and risk management. By integrating multiple kernels within the SVC model, the final map provided a more robust and comprehensive evaluation of landslide risk across the study area. These maps can serve as invaluable resources for stakeholders, including policymakers, urban planners, and disaster-management authorities, thus facilitating proactive measures for mitigating landslide-related risks.

– A comprehensive feature importance analysis was conducted to evaluate the impacts of various factors on landslide occurrences, thus providing essential insights into those elements that are most strongly associated with landslide-susceptibility. This analysis was performed for each kernel of the SVC model by categorizing the features into three levels of importance: low, moderate, and high. This structured approach enhances our understanding of how each feature contributes to landslide prediction. The findings indicated that, for SVC_linear, rainfall and soil type were identified as high-importance features, while for SVC_poly, aspect was classified under the list of high-importance features. In SVC_rbf, drainage density, NDVI, rainfall, and soil type were classified as high-importance features. The results highlighted several critical features, including rainfall, drainage density, the NDVI, soil type, and aspect. These findings emphasized the importance of focusing on these factors in those regions that are prone to landslides. To effectively mitigate landslide risks, it is imperative to implement measures such as enhanced rainfall monitoring systems, improved drainage infrastructures, vegetation restoration initiatives, and comprehensive studies of soil properties. By addressing these key factors, we can develop targeted mitigation strategies that significantly reduce the frequency and impacts of landslides, thereby promoting safety and sustainability in vulnerable areas.

Funding

This research received no specific grant from any funding agency in the public, commercial, nor not-for-profit sectors.

Declaration of Competing Interests

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work that was reported in this paper.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Use of Generative AI and AI-Assisted Technologies

Generative AI tools like ChatGPT and Quillbot were used in this research solely for grammatical corrections. No content was generated through generative AI in this study.

Acknowledgement

The author would like to express his heartfelt gratitude to the faculty and staff of the Department of Civil Engineering and the Department of Computer Science at Maulana Azad National Institute of Technology, Bhopal, India. Their unwavering support and guidance have been instrumental throughout the research process. The author is especially appreciative of the invaluable expertise that was shared by the dedicated faculty members, whose insights and constructive feedback greatly enhanced the quality of this study. It is through their commitment to academic excellence and mentorship that this work was able to reach its full potential.

References

[1] Kjekstad O., Highland L.: *Economic and social impacts of landslides*, [in:] Sassa K., Canuti P. (eds.), *Landslides – Disaster Risk Reduction*, Springer, Berlin, Heidelberg, pp. 573–587. https://doi.org/10.1007/978-3-540-69970-5_30.

[2] Tien Bui D., Tuan T.A., Klempe H., Pradhan B., Revhaug I.: *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, vol. 13(2), 2016, pp. 361–378. https://doi.org/10.1007/s10346-015-0557-6.

[3] Gerrard J.: *The landslide hazard in the Himalayas: Geological control and human action*, [in:] Morisawa M. (ed.), *Geomorphology and Natural Hazards, Proceedings of the 25th Binghamton Symposium in Geomorphology, Held September 24–25, 1994 at SUNY, Binghamton, USA*, Elsevier, 1994, pp. 221–230. https://doi.org/10.1016/B978-0-444-82012-9.50019-0.

[4] Bera A., Mukhopadhyay B.P., Das D.: *Landslide hazard zonation mapping using multi-criteria analysis with the help of GIS techniques: A case study from Eastern Himalayas, Namchi, South Sikkim.* Natural Hazards, vol. 96(2), 2019, pp. 935–959. https://doi.org/10.1007/s11069-019-03580-w.

[5] Wang Y., Sun D., Wen H., Zhang H., Zhang F.: *Comparison of random forest model and frequency ratio model for landslide-susceptibility mapping (LSM) in Yunyang County (Chongqing, China).* International Journal of Environmental Research and Public Health, vol. 17(12), 2020, 4206. https://doi.org/10.3390/ijerph17124206.

[6] Vakhshoori V., Zare M.: *Landslide-susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods.* Geomatics, Natural Hazards and Risk, vol. 7(5), 2016, pp. 1731–1752. https://doi.org/10.1080/19475705.2016.1144655.

[7] Ozdemir A., Altural T.: *A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide-susceptibility mapping: Sultan Mountains, SW Turkey.* Journal of Asian Earth Sciences, vol. 64, 2013, pp. 180–197. https://doi.org/10.1016/j.jseaes.2012.12.014.

[8] Wu C.H.: *Landslide-susceptibility mapping by using landslide ratio-based logistic regression: A case study in the southern Taiwan.* Journal of Mountain Science, vol. 12(3), 2015, pp. 721–736. https://doi.org/10.1007/s11629-014-3416-3.

[9] Althuwaynee O.F., Pradhan B., Ahmad N.: *Landslide-susceptibility mapping using decision-tree based CHi-squared automatic interaction detection (CHAID) and logistic regression (LR) integration.* IOP Conference Series: Earth and Environmental Science, vol. 20(1), 2014, 012032. https://doi.org/10.1088/1755-1315/20/1/012032.

[10] Rasyid A.R., Bhandary N.P., Yatabe R.: *Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia.* Geoenvironmental Disasters, vol. 3, 2016, 19. https://doi.org/10.1186/s40677-016-0053-x.

[11] Cervi F., Berti M., Borgatti L., Ronchetti F., Manenti F., Corsini A.: *Comparing predictive capability of statistical and deterministic methods for landslide-susceptibility mapping: a case study in the northern Apennines (Reggio Emilia Province, Italy).* Landslides, vol. 7(4), 2010, pp. 433–444. https://doi.org/10.1007/s10346-010-0207-y.

[12] Tang R.X., Yan E.C., Wen T., Yin X.M., Tang W.: *Comparison of logistic regression, information value, and comprehensive evaluating model for landslide-susceptibility mapping.* Sustainability, vol. 13(7), 2021, 3803. https://doi.org/10.3390/su13073803.

[13] Arabameri A., Rezaei K., Pourghasemi H.R., Lee S., Yamani M.: *GIS-based gully erosion susceptibility mapping: A comparison among three data-driven models and AHP knowledge-based technique.* Environmental Earth Sciences, vol. 77(17), 2018, 628. https://doi.org/10.1007/s12665-018-7808-5.

[14] Khatun M., Hossain A.S., Sayem H.M., Moniruzzaman M., Ahmed Z., Rahaman K.R.: *Landslide-susceptibility mapping using weighted-overlay approach in Rangamati, Bangladesh.* Earth Systems and Environment, vol. 7(1), 2023, pp. 223–235. https://doi.org/10.1007/s41748-022-00312-2.

[15] Bopche L., Rege P.P.: *Landslide-susceptibility mapping: An integrated approach using geographic information value, remote sensing, and weight of evidence method.* Geotechnical and Geological Engineering, vol. 40(6), 2022, pp. 2935–2947. https://doi.org/10.1007/s10706-022-02070-4.

[16] Pal S.C., Chowdhuri I.: *GIS-based spatial prediction of landslide-susceptibility using frequency ratio model of Lachung River basin, North Sikkim, India.* SN Applied Sciences, vol. 1(5), 2019, 416. https://doi.org/10.1007/s42452-019-0422-7.

[17] Wang L.J., Guo M., Sawada K., Lin J., Zhang J.: *Landslide-susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models.* Catena, vol. 135, 2015, pp. 271–282. https://doi.org/10.1016/j.catena.2015.08.007.

[18] Chen W., Xie X., Wang J., Pradhan B., Hong H., Bui D.T., Duan Z., Ma J.: *A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide-susceptibility.* Catena, vol. 151, 2017, pp. 147–160. https://doi.org/10.1016/j.catena.2016.11.032.

[19] Pourghasemi H.R., Pradhan B., Gokceoglu C.: *Application of fuzzy logic and analytical hierarchy process (AHP) to landslide-susceptibility mapping at Haraz watershed, Iran.* Natural Hazards, vol. 63, 2012, pp. 965–996. https://doi.org/10.1007/s11069-012-0217-2.

[20] Felicísimo Á.M., Cuartero A., Remondo J., Quirós E.: *Mapping landslide-susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study.* Landslides, vol.10, 2013, pp. 175–189. https://doi.org/10.1007/s10346-012-0320-1.

[21] Hong H., Pradhan B., Xu C., Bui D.T.: *Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena*, vol. 133, 2015, pp. 266–281. https://doi.org/10.1016/j.catena.2015.05.019.

[22] Colkesen I., Sahin E.K., Kavzoglu T.: *Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression.* Journal of African Earth Sciences, vol. 118, 2016, pp. 53–64. https://doi.org/10.1016/j.jafrearsci.2016.02.019.

[23] Wang Y., Fang Z., Wang M., Peng L., Hong H.: *Comparative study of landslide-susceptibility mapping with different recurrent neural networks.* Computers & Geosciences, vol. 138, 2020, 104445. https://doi.org/10.1016/j.cageo.2020.104445.

[24] Wang Y., Fang Z., Hong H.: *Comparison of convolutional neural networks for landslide-susceptibility mapping in Yanshan County, China.* Science of the Total Environment, vol. 666, 2019, pp. 975–993. https://doi.org/10.1016/j.scitotenv.2019.02.263.

[25] Lee S., Ryu J.H., Lee M.J., Won J.S.: *The application of artificial neural networks to landslide-susceptibility mapping at Janghung, Korea.* Mathematical Geology, vol. 38, 2006, pp. 199–220. https://doi.org/10.1007/s11004-005-9012-x.

[26] Tsangaratos P., Benardos A.: *Estimating landslide-susceptibility through an artificial neural network classifier.* Natural Hazards, vol. 74, 2014, pp. 1489–1516. https://doi.org/10.1007/s11069-014-1245-x.

[27] Lee S., Ryu J.H., Won J.S., Park H.J.: *Determination and application of the weights for landslide-susceptibility mapping using an artificial neural network.* Engineering Geology, vol. 71(3–4), 2004, pp. 289–302. https://doi.org/10.1016/S0013-7952(03)00142-X.

[28] Cortes C., Vapnik V.: *Support-vector networks.* Machine Learning, vol. 20, 1995, pp. 273–297. https://doi.org/10.1007/BF00994018.

[29] Boser B.E., Guyon I.M., Vapnik V.N.: *A training algorithm for optimal margin classifiers,* [in:] *COLT '92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Association for Computing Machinery, New York 1992, pp. 144–152. https://doi.org/10.1145/130385.130401.