Application of Data-Driven Techniques For Landslide Susceptibility Prediction Along An Earthquake Affected Road Section in Kashmir Himalaya

Syed Ahsan Hussain Gardezi
Geological Survey of Pakistan
Nadeem Ahmad Usmani (n.usmani145@gmail.com)
Geological Survey of Pakistan https://orcid.org/0000-0002-6340-0074
Xiao-qing Chen
Institute of Mountain Hazards and Environment Chinese Academy of Sciences
Nawaz Ikram
The University of Azad Jammu & Kashmir
Sajjad Ahmad
Geological Survey of Pakistan
Saad Wani
University of Azad Jammu and Kashmir

Research Article

Keywords: Landslide susceptibility, machine learning algorithm, statistical model, heuristic model, Kashmir Himalaya

Posted Date: December 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1105939/v1

APPLICATION OF DATA-DRIVEN TECHNIQUES FOR LANDSLIDE SUSCEPTIBILITY PREDICTION ALONG AN EARTHQUAKE AFFECTED ROAD SECTION IN KASHMIR HIMALAYA

Syed Ahsan Hussain Gardezi[1], Nadeem Ahmad Usmani*[,1,2,3], Xiao-qing Chen[2,3], Nawaz Ikram[4], Sajjad Ahmad[1], Saad Wani[4]

[1]Geological Survey of Pakistan, Islamabad, 45320, Pakistan.
[2]Key Laboratory of Mountain Hazards and Earth Surface Process, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, 610041, China.
[3]University of Chinese Academy of Sciences, Beijing, 100049, China.
[4]Institute of Geology, University of Azad Jammu and Kashmir, Muzaffarabad, 13100, Pakistan.

*Corresponding Author,
Email: n.usmani145@gmail.com

Conflicts of interest:

All authors have no conflicts of interest to disclose that are relevant to the content of this article.

Highlights:

1. Landslide susceptibility mapping along earthquake affected road-section in Kashmir Himalaya
2. Thirteen causative factors involved in Prediction of landslide susceptibility using six data-driven models
3. Machine learning models outperformed heuristic and statistical models
4. Distribution of predicted landslides was common in the proximity of road/faults and deeply dissected valley slopes

Abstract

The interaction of seismic events with geo-environmental conditions and anthropogenic activities may exacerbate the risk of landslide hazard in a mountainous region. As an example of this, 2005 Kashmir earthquake triggered a large number of shallow to deep slope failures, which was further intensified in following years by human activities notably along road networks, posing a long-term hazard. Hence, this study was planned to evaluate the effectiveness of landslide susceptibility prediction along earthquake affected road-section of Neelum Highway using six different data-driven models. We applied analytical hierarchy process as heuristic approach, weight of evidence and index of entropy as statistical models and multi-layer perceptron, support vector machine and binary logistic regression (BLR) as machine learning models. Initially, 224 landslides locations were marked through field surveys to prepare landslide inventory which was further randomly divided into training (70%) and testing (30%) datasets. Then, 13 landslide causative factors (LCFs) were extracted from geo-spatial database and analysed by measuring collinearity among factors and assessing their contribution in landslide occurrence using different feature selection methods for inclusion in susceptibility modelling. Thereafter, six employed models were trained to produced landslide susceptibility maps of investigated road-section. Finally, the area under receiver operating characteristics (AU-ROC) curve and various statistical measures were applied to validate and compare the performance of modeled landslide susceptibility. The results revealed that no collinearity issue exists among all 13 LCFs, and all six models exhibited satisfying performance in predicting landslide susceptibility of study area. However, BLR model have produced most promising and optimum results as compared to other models with AU-ROC (0.881), Matthew's correlation coefficient (0.609), Kappa coefficient (0.604), accuracy (0.797) and F-score (0.787). The outcomes of this study can be used as pertinent guide for preventing and managing the landslide disaster risk along Neelum Highway and beyond.

Keywords: Landslide susceptibility; machine learning algorithm; statistical model; heuristic model; Kashmir Himalaya

1 Introduction

Landslide is a kind of complex geohazard in the mountainous terrains around the world that imposes an enormous threat to population, sources of livelihood and infrastructure. It occurs as a result of the interplay of geological events, geo-environmental factors and anthropogenic activities (Aleotti and Chowdhury, 1999; Guzzetti, 2005). Globally, the Himalaya is one of the most favourable and dynamic mountainous terrain for destructive landslide hazard where combine impact of periodic seismicity, monsoon rains and human activities along unstable slopes often aggravate the phenomena of landsliding (Gerrard, 1994; Rusk et al., 2022). The region of Kashmir Himalaya is falling in seismically active zone of western Himalayan belt, had faced devastating seismic events in the past with most recent catastrophic earthquake of 2005 (Gardezi et al., 2021; Sana and Nath, 2017). This earthquake event and aftershocks had triggered more than two thousands of landslides over an affected area of around 7500 square kilometres. The devastation had badly affected the major components of Neelum Highway and resulted in severe damaging and blocking of main route even three months after the main event (Basharat et al., 2014a; Rahman et al., 2014). This mega event and post-earthquake shaking highly deformed the rocks and significantly increased vulnerability of slopes for landsliding which leads to the development and reactivation of landslides upon subsequent interaction with other extrinsic variables (Khattak et al., 2010). Therefore, it is crucial to assess the possible zones of slope failures in this region using modern practices to formulate effective landslide mitigation strategies. For this purpose, the landslide susceptibility mapping is considered as a primary component in landslide hazards and risk assessment in past decades (Huang et al., 2020).

Landslide susceptibility represents the spatial probability of landslides in a particular region based on local geo-environmental conditions and causative factors. Over the years, the number of studies were performed by several investigators e.g., (Achour and Pourghasemi, 2020;

Aditian et al., 2018; Chen et al., 2019; Guzzetti, 2005; Hong et al., 2020; Huang et al., 2020; Merghadi et al., 2020; Saha and Saha, 2020; Wang et al., 2021; Wang et al., 2016; Zhang et al., 2018) to assess the landslide susceptibility using different modelling approach, which can be generally categorised into deterministic and data-driven models. In general, the deterministic models calculate the quantitative stability factor of slope in certain area to predict landslides events. These models are usually applied to a local area having simple landslide types and homogamous geologic and geomorphic conditions, however the hazard complexity and extensive data acquisition make it difficult to analyse over large scale in heterogeneous conditions (Ciurleo et al., 2017). In contrast, data-driven models exhibit more reliable results with high accuracy in predicting regional landslide susceptibility and can be categorized into heuristic, statistical and machine learning models (Huang et al., 2020). The heuristic approach can execute landslide susceptibility prediction (LSP) by measuring the weight values for each conditioning factor based on expert opinion derived from the field knowledge using rating techniques, mainly include analytical hierarchy process (AHP) and weighted linear combination (WLC) (Shahabi and Hashim, 2015). On the other hand, statistical approach predicts the probability of landsliding through numerical relationship between observed landslides and causative factors (Erener and Düzgün, 2012). The statistical techniques such as certainty factor (Hong et al., 2017), evidential belief function (Ding et al., 2017), frequency ratio (Li et al., 2017; Yalcin et al., 2011), generalized linear models (Goetz et al., 2011; Huang et al., 2020), information value method (Chen et al., 2014; Du et al., 2017), index of entropy (IoE) and weight of evidence (WoE) (Liu and Duan, 2018; Regmi et al., 2014) have been successfully applied in various studies to produce landslide susceptibility maps (LSM) with significant model accuracy. Similarly in last few years, machine learning (ML) models are rapidly evolving in predicting landslide susceptibility due to high precision rate and adequate handling of complex data. The commonly used ML methods in LSP include AdaBoost (Hong

et al., 2018), binary logistic regression (Huang et al., 2020; Pradhan and Jebur, 2017), decision tree (Dou et al., 2019), gradient boosting (Chen et al., 2020; Sahin, 2020), K-nearest neighbors (Pradhan and Jebur, 2017), kernal logistic regression (Aditian et al., 2018; Park et al., 2013), multilayer perceptron (MLP) (Bui et al., 2020; Zare et al., 2013), naïve bayes (Chen et al., 2017; Tien Bui et al., 2012), neuro fuzzy (Pradhan, 2013), random forest (Chen et al., 2020; Dou et al., 2019) and support vector machine (SVM) (Bui et al., 2020; Pradhan, 2013).

Although, a number of approaches have been proposed for producing LSMs in existing literature, but still no consensus developed among scholars around the globe that which model is more reliable and best suited for predicting landslide susceptible areas in different terrains. However, researchers recommend that these models should be applied and tested in different geo-environmental conditions, where landslide activities are governed by multiple processes to validate the quality and accuracy of model used (Pourghasemi and Rahmati, 2018). Moreover, few studies suggest the practical application of different models and their comparative analysis in a similar area should be investigated to obtain a robust model for LSP. Therefore, in light of foregoing consideration, it is indispensable to implement and compare popular data-driven techniques in a given area for better understanding of landslide susceptibility prediction.

As the most disastrous and tectonically active region of Kashmir Himalaya, Muzaffarabad-Neelum trunk road was chosen as example for present study. This road-section experienced a devastating 7.6 M_w earthquake that triggered enormous landslides along natural and cut slopes with subsequent blockage of accessed roads (Kamp et al., 2008; Owen et al., 2008). The main seismic event and its aftershocks developed extensive fissures, cracks, lateral and back scarps in surrounding slopes of this road-section, which resulted in frequent and progressive slope failure specially in extreme weather conditions. Hence, it is essential to understand and recognize the persistent patterns of landslide hazard for mitigating future slope failures in this

region. Most of studies have been conducted to identify the occurrence, size, spatial distribution and pattern of landslides associated with 2005 Kashmir earthquake e.g., (Ahmed et al., 2021; Basharat et al., 2014a; Basharat et al., 2014b; Basharat et al., 2016; Jadoon et al., 2015; Kamp et al., 2008; Khan et al., 2013; Khattak et al., 2010; Owen et al., 2008; Saba et al., 2010; Shafique, 2020; Shafique et al., 2016). However, only few studies were focused on landslide susceptibility mapping using conventional quantitative techniques such as Basharat et al. (2016) and Kamp et al. (2008) undertook LSP for earthquake affected areas based on semi-quantitative approach (AHP). Recently, Ahmed et al. (2021) demarcated the landslide prone areas along a small portion of Neelum road using WOE method and investigated the geo-technical aspect of specific landslides. Though these studies demonstrate landslide hazard maps which were modulated by limited number of landslide conditioning factors, whereas a comprehensive study on spatial prediction of landslide susceptibility using popular machine learning models is still lacking for this region. We therefore undertook an extensive analysis of machine learning techniques (MLP; SVM; BLR) and their comparison with heuristic (AHP) and traditional statistical methods (IoE; WoE) using a broad range of causative variables to evaluate a robust model for accurate LSP along earthquake affected road-section in Kashmir Himalaya.

2 Study area

The study area lies in the western part of main landmass of the State of Jammu and Kashmir and is approximately situated between the longitudes 73°26' E to 73°56' E and latitudes 34°21' N to 34°40' N. It is positioned along the main drainage of Neelum River and covers an area of ~287.97 sq.km (Fig. 1c, d). The Athmuqam city is located at eastern, Nauseri in the central while Muzaffarabad city falls in the western part of the study area connected by a major trunk road. This road is the lifeline for more than 0.3 million inhabitants of Neelum valley. Moreover, it has a significant strategic importance as it is located along Pakistan–India Line of Control.

Topographically, the study area dominantly characterized by hilly terrain with elevation ranges from 632 m to 2857 m above sea level (asl). The terrain in northern and central regions of the study area are relatively higher with altitude exceeding ~1100 m asl as compared to western region which have mean elevation of ~800 m asl (Fig. 1d). The settlements in the study area are widely dispersed along gentle slopes, river and streams terraces and valley floors increasing pressure on the land and environment due to high population density and small land holdings. The climate of the region is humid, humid tropical and subtropical highland with average annual rainfall of 1400 mm. Mean maximum and minimum temperatures in the area during summer are about 20-34°C and in winter 2-18°C, respectively (Bashir and Hanif, 2018).

2.1 Geomorphological setting

The study area being a western part of Lesser and Sub-Himalayas is classified into three geomorphological zones i.e., mountainous terrain, the lower hills of valley slopes and foot slopes of fluvial sediments along the major tributaries. The valley floor in the northern part is deeply incised by Neelum river and its feeding tributaries having a relief of about 2200 m. The rapid flow of the river in this region with an average discharge of nearly 230 m^3/s has resulted in fluvial incision forming a steeper low valley with slopes exceeding 50°. This landscape reflects a high relief energy resulting a V- shaped valley with high erosion rate. The back slopes above these valley slopes are less steep with average slopes of about 15–30° (Kamp et al., 2008). Talus or slided-mass, fluvial deposits, terraces, alluvial fans and glaciofluvial deposits are major geomorphic accumulations observed in the study area.

2.2 Geological and tectonic settings

Geologically, the area is comprised of sedimentary, metamorphic and igneous rocks which ranges from Precambrian to Miocene. The lithostratigraphic units exposed in the northern part of the study area are Precambrian Salkhala Formation comprising of talc-quarzitic schists, graphitic schist with subordinate marble and Cambrian Jura Granite of gneissic and porphyritic

type with doloritic intrusions at certain places. The Panjal Formation consist of greenstone, andesitic flows with bedded tuff, and meta-carbonates with agglomeratic slate, phyllite and mica schist exposed in the central part. The Tertiary fluvio-clastic deposits of Murre Formation containing sandstone, mudstone and siltstones formed as a consequence of Himalayan orogeny (Bossart et al., 1988) occupies the central and southern part of the study area. The Paleocene-Eocene carbonate deposits are unconformably overlies along the cherty dolomitic limestone of Cambrian Abbottabad Formation near Muzaffarabad. The quaternary deposits of fluvial origin are mostly exposed along the lower slopes and Terraces (Hussain and Khan, 1996; Shah, 2009).

The study area lies in the core and limbs of prominent structure i.e., Hazara–Kashmir Syntaxis (HKS), which implicates the region as complex tectonic zone(Calkins et al., 1975; Hussain et al., 2020). Tectonically, the area is sub-divided into three units: a) the youngest Tertiary Molasse deposit in the core of HKS surrounded by the Main Boundary Thrust (MBT) and Pre-molasse sequences, b) the tholeiitic-alkaline volcanic and agglomeratic slates, extending as a linear belt between the MBT and the Panjal Thrust, c) the oldest Precambrian meta-sediments and Cambrian granites exposed on the eastern and western limbs of HKS. The MBT in the study area constitute a suture along which remnants of a Paleozoic ocean floor sediment, the Panjal Volcanics and the agglomeratic slates have been thrusted over the Murree Formation (Baig, 2006; Hussain et al., 2009) (Fig. 1b, c). The western limb of the HKS is more complicated by the emergence of Balakot-Bagh Fault (BBF) which emplaced the Cambrian dolomite of Muzaffarabad Formation over Miocene Murree Formation. Along BBF, seismic activity of 2005 Earthquake had generated a mean vertical displacement of ~4.3 m in the central segment of the fault near Chela Bandi, Muzaffarabad (Kaneda et al., 2008). This event created many cracks on the hillslope resulted in slope failure of bed rock as well as colluvial deposits along major roads, natural slopes and other man-made slope cuts (Owen et al., 2008).

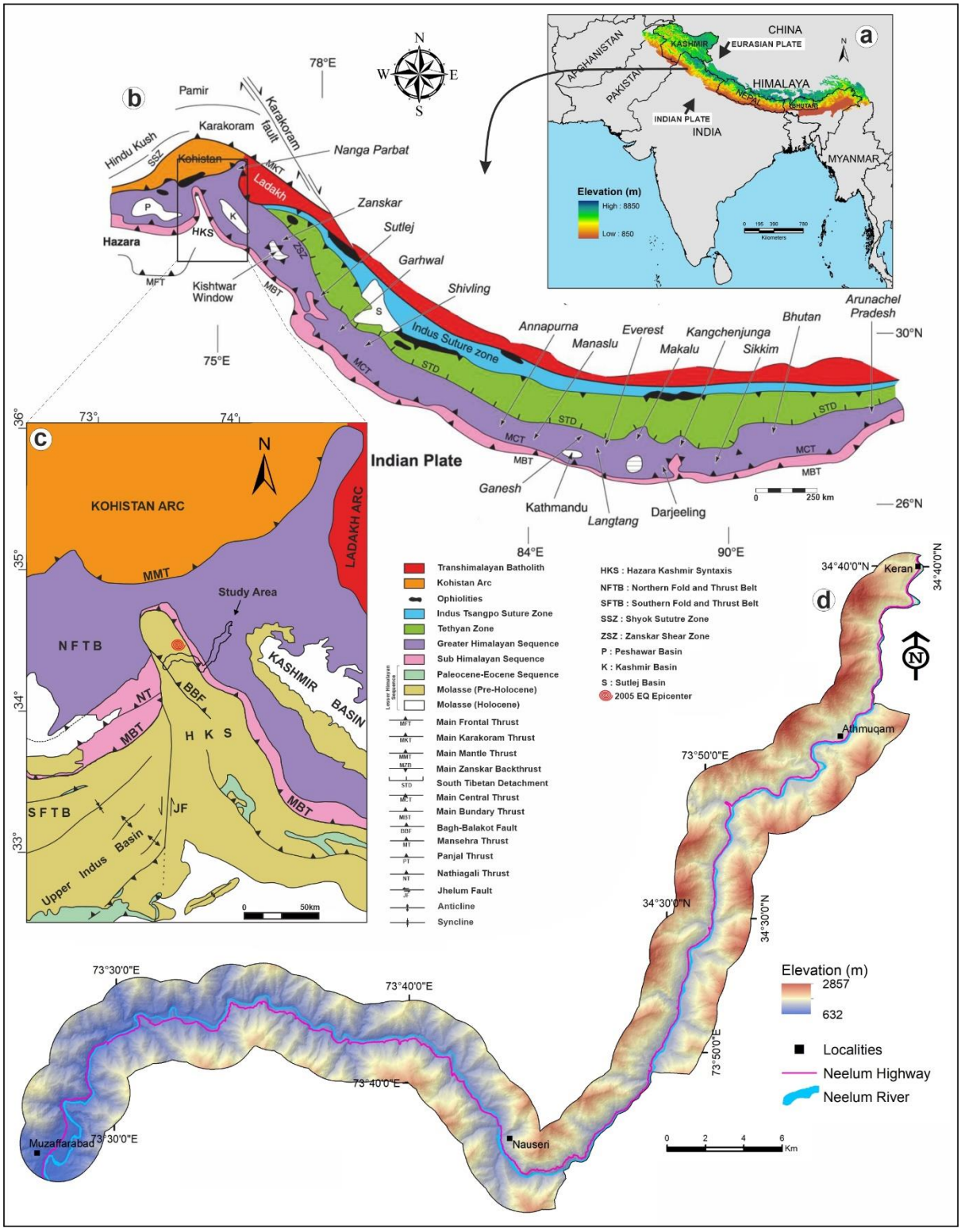

Fig. 1 Location map of study area a) geographical position of Himalaya; b) geological map of Himalaya (Searle and Treloar, 2019); c) extent map indicating structural setup of NW and Kashmir Himalaya (Baig, 2006) ;d) route map of Neelum Highway

3 Dataset preparation

The landslides susceptibility modelling in a given area involves the collection of data on landslide occurrence, geospatial database construction on landslide conditioning factors, and

probability modelling for landslide susceptible areas based on relationship between landslide events and their triggering factors followed by validation of produced models (Guzzetti, 2005; Regmi et al., 2014). For this purpose, the dataset we used to construct different thematic layers includes ALOS PALSAR DEM with 12.5 m resolution, geological maps (GSP), published reports and research articles, Landsat-8 Images (USGS), Google Earth imageries, rainfall data (TRMM; Pakistan Metrological Department) and detailed field mapping.

3.1 Landslide inventory mapping

The preparation of accurate landslide inventory is an important step for investigating the possible landslide hazards and susceptibility mapping (Guzzetti et al., 2012). These inventory maps demonstrate information about the landslide locations, its dimensions and types of mass movements that have triggered in the past. In present study, the landslide distribution has been extracted from previous literatures (Basharat et al., 2014a; Kamp et al., 2008; Owen et al., 2008), historical records of GSP, and visual interpretation of Google Earth images (2003-2020). The extensive field survey was also carried out during March-April 2021 for validation, precise demarcation of landslides area and their characteristics. For this purpose, the landslides dataset recorded systematically during field observation using GPS based Field MOVE Clino (Android App) and finally converted into vector database in GIS for preparation of landslide inventory map. A total of 224 landslide events were recorded during this study and are illustrated in (Fig. 2).

The landslide distribution analysed during fieldwork reveals that geo-environmental and anthropogenic factors are controlling mass movements in most part of the study area, whereas co-seismic landslides are mostly concentrated along the proximity of BBF particularly around the north-western portion of hanging wall slopes of fault trace. Furthermore, the majority of landslides are rock falls/slides and debris slides, while the types of movement are translational, rotational and complex (Fig. 3). The clustered failures were mostly shallow involving bedrock

to superficial accumulations, whereas size of the materials ranges from few m^3 to thousands of m^3. Based on landslides inventory mapping, the study area can broadly be categorized into three sections: a) northern section is a narrow dissected valley characterized by rockslide, rockfall and debris slide of steeply jointed rocks of Precambrian metamorphic and Cambrian granites; b) comparatively broader valley in the central section having less steep slopes (<50°) triggering mixed landslides mostly associated with road cuts and artificial water channels in Tertiary molasses and Quaternary fluvial terraces; c) in the southern section, rockfalls-mixed landslides developed in Cambrian dolomitic limestone and Paleocene-Eocene marine sediments mostly clustered along the proximity of active BBF.

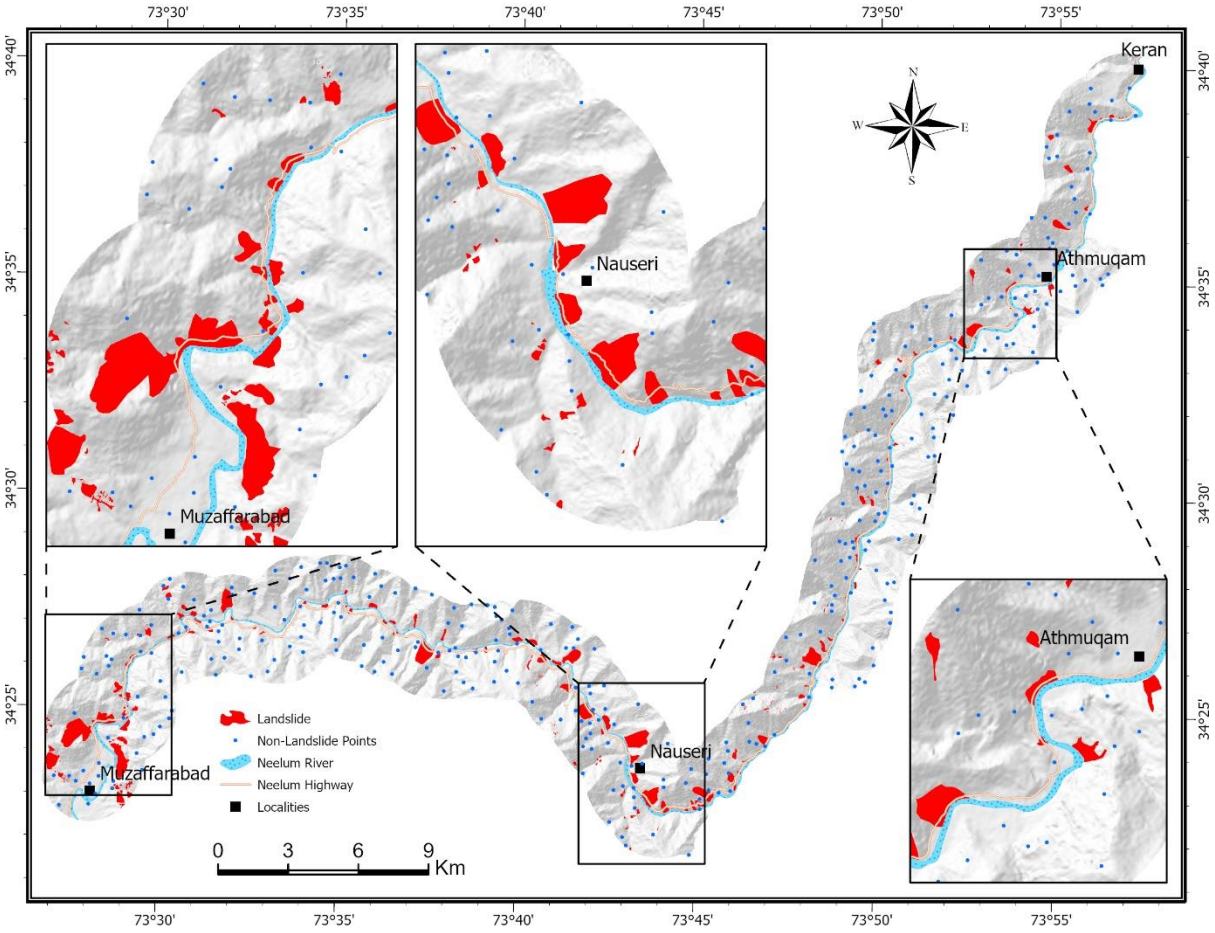

Fig. 2 Landslide inventory map of study area

In current study, the compiled landslide inventory was converted as landslides point data based on sampling strategy specified by (Dou et al., 2020), whereas equal number of non-landslides points were randomly selected from stable areas. This dataset was then converted

into training (70%) and testing (30%) data using ArcGIS “Subset” tool for modelling landslide susceptibly and its validation.

Fig. 3 Field photographs a) Landslide in Murre Formation near Deolian village; b) showing several episodes of slope failure in Salkhala Formation near Barian area; c) Dhani Landslide triggered in the hanging wall of BBF; d) multiple phases of Panjgran Landslide caused by combine interaction of road- and river-cutting and fragile lithology; e) Shahkot landslide initiated due to the water seepage from irrigation channel on upper half and toe-cutting

3.2 Landslide causative factors

The selection of landslide causative factors (LCFs) generally rely on the available data, scale and nature of study area as there is no set criteria or universal agreement until now (Nefeslioglu et al., 2008). Our study area lies in the Himalayan region where active tectonics, geo-environmental variables and anthropogenic activities are main contributors of mass movements. Therefore, thirteen parameters consisting of geological factors (lithology and proximity to fault), geomorphic factors (slope, aspect, curvature, elevation, topographic wetness index (TWI), proximity to drainage and lineaments density), environmental factors (Normalized Difference Vegetation Index (NDVI), landuse and precipitation) and anthropogenic factor (proximity to roads) were selected as LCFs and were standardized to a common cell size (12.5 x 12.5 m).

3.2.1 Geological factors

The fragile lithology, sheared and jointed rocks affects the stability of materials which often leads to landslide event along steep slopes (Basharat et al., 2014a; Kamp et al., 2008). Therefore, in present study lithology and distance to fault were taken as input variables for landslide susceptibility modelling.

3.2.1.1 Lithology

Lithology is classified in terms of rock and soil composition, whereas, its mechanical properties and composition controls the permeability and strength of slope forming materials, thus considered as an important parameter in landslide susceptibility mapping (Ayalew and Yamagishi, 2005; Yalcin et al., 2011). The information on the spatial distribution of lithological formations were identified from the geological maps of GSP (Hussain and Khan, 1996; Hussain et al., 2006; Hussain et al., 2009). Eleven geological formations are exposed, which comprised of lithological units derived from igneous, metamorphic and sedimentary origin such as granite, doleritic intrusions, schist, slates, phyllites, marble, basaltic volcanic flows, dolomitic

limestone, cherty nodular limestones, marine shales, sandstone, siltstone and mudstone with quaternary fluvial terrace, alluvial fans and colluvial deposits at certain places (Fig. 4h).

3.2.1.2 Proximity to faults

The faults are categorized as zone weakness and are responsible for rocks fracturing and jointing which controls the failure mechanism in the form of rock sliding and wedge sliding. The tectonic stresses in the proximity to faults controls the permeability of terrain and affect the slope stability is considered as the important influencing parameter for slope failures in landslide susceptibility (Dou et al., 2020; Pradhan and Jebur, 2017). The study area being part of North-eastern Himalayas of Pakistan is recognized as the zone of high seismicity with active thrust faulting often contributing seismic hazard accompanied by lethal landsliding (Gardezi et al., 2021). The most recent seismic event of 2005 had produced thousands of landslides with extensive cracks, fissures and bulging slopes in the region (Basharat et al., 2014a; Kamp et al., 2008; Owen et al., 2008). The field observations recorded in the study area revealed that rocks fracturing was common over a proximity of 500 m from fault, whereas the intense zones of fractures were seen in the range of 0-150 m. Therefore, for understanding the relationship between fault presence and landslide occurrence, five buffer zones were created along the major fault lines in ArcGIS "Euclidian distance" tool and classified accordingly: <150 m, 150–300 m, 300-450 and >500 m (Fig. 4i).

3.2.2 Geomorphic factors

Geomorphological parameters are considered as major influencing terrain factor in landslide occurrence (Pourghasemi and Rahmati, 2018). The geomorphic parameters we used in this study are briefly described below:

3.2.2.1 Slope angle

Slope angle plays a vital role in contributing landslide activity and is considered an important causative factor in susceptibility modelling of landslide hazards (Huang et al., 2020; Wang et al., 2020a). The common slope failure generally depends on the degree of cohesion in slope material and inclined angle where gravity exerts a sliding force on slope mass (Pourghasemi et al., 2018). In our case, the slope gradient ranges from 0-80.27° and were reclassified into seven classes based on landslide distribution analysed during field study. The slope angle categorized with10° interval as: <10°, 10–20°, 20–30°, 30–40°, 40–50°, 50–60°and <60° (Fig. 4a).

3.2.2.2 Aspect

The aspect of slope is generally described as slope facing direction having significant impact on several processes (vegetation distribution, discontinuities patterns, weathering and moisture retention) which can directly or indirectly effects the landslide phenomena (Shahabi and Hashim, 2015) . The environmental conditions such as snow melt precipitation pattern and sunlight exposure also influence the slope stability with subsequent changes along particular slope direction (Ayalew and Yamagishi, 2005). The slope aspect in this study was classified as N, NE, E, SE, S, SW, W, NW and flat (Fig. 4b).

3.2.2.3 Curvature

Slope curvature is one of the important geomorphological factors widely used as LCFs in landslide susceptibility mapping (Dou et al., 2015). It represents the concavity, flatness and convexity of terrain surfaces which stimulates the flow velocity and divergence of runoff water and controls the occurrences of landslide (Bera et al., 2019). The curvature values for current study were reclassified as flat (zero curvature), concave (+tive curvature) and convex (-tive curvature) (Fig. 4c).

3.2.2.4 Elevation

The altitude has a strong relationship with mass movement and is considered as primary risk factor that controls the secondary causative variables such as weather conditions, vegetation, human activities and other conditions (Aditian et al., 2018). Various landslide studies have shown that slope failures are scarce at higher altitude (Pourghasemi et al., 2018). In current study, the elevation data obtained from DEM was sub-divided into five classes with an average interval of 500m (Fig. 4d).

3.2.2.5 TWI

TWI indicates the topographic control on local hydrological conditions, evaluated through soil saturated zones of an area with respect to catchment locations (Wang et al., 2020a). The degree of water distribution in a soil increases the pore pressure which affect the soil, rock and vegetation condition on upslope resulting in slope failure (Sun et al., 2020). According to Beven and Kirkby (1979), it is defined as:

$$TWI = ln\left(\frac{a}{tan\,\beta}\right) \quad (1)$$

where α is expressed as the flow accumulation in certain point and β is measured as the local degree slope (radian). In present study, the TWI values were calculated through DEM using Arc GIS operations and were reclassified as: <5, 5-9 and >9 (Fig. 4l).

3.2.2.6 Proximity to drainage

The proximity to drainage is considered as an important landslide causative factor, influencing the slope stability in terms of erosion along the toes of slopes and by increasing the water saturation level in lower part of the slope (Bera et al., 2019). This effect become adverse when water laden drainage borders the steep slopes and sliding zones because the fluvial incision and lateral cutting gradually favours the condition for mass movement (Sun et al., 2020). In the study area, the proximity to drainage was calculated through "Euclidean distance"

tool in GIS and was classified into 4 classes with an interval of 100 m: 0-100 m, 100-200 m, 200-300 m and > 300 (Fig. 4k).

3.2.2.7 Lineaments density

The lineaments pattern in a specific area can be expressed as linear topographic signatures controlled by geomorphic and tectonic processes. These features may affect the stability of slopes and increases the probability of slope failures (Ayalew and Yamagishi, 2005). The lineament patterns were prepared through visual interpretation of Landsat 8 OLI imagery by using PCI Geomatica software. The extracted lineaments were processed in ArcGIS by Euclidean density function to generate the lineaments density map with five classes: 0-0.45 km/km^2, 0.45-0.9 km/km^2, 0.9-1.35 km/km^2, 1.35-1.9 km/km^2 and 1.9-3.5 km/km^2 (Fig. 4g).

3.2.3 Environmental factors

3.2.3.1 NDVI

The NDVI is measured as degree of vegetation distributed in a region and is closely associated with water flow and penetration and soil erosion which influence the landslide activity along slope surfaces (Huang et al., 2020). Therefore, it has been effectively utilized as a crucial parameter in landslide prediction modelling. In current study, the values of NDVI were computed using Landsat 8 OLI multi spectral imagery of 15th Sep, 2019 (https://earthexplorer.usgs.gov) by following formula:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (2)$$

where R is computed as the red spectral band and NIR is near infrared band of electromagnetic spectrum (Tucker, 1979). These values range from -0.15 to 0.56 and were categorized as <0.1, 0.1-0.3 and >0.3 (Fig. 4e).

3.2.3.2 Landuse

The landuse is frequently used contributing factor in susceptibility modelling for landslide especially in the mountainous terrain where natural process effect the slope stability of a region (Pourghasemi et al., 2018). The changes in land cover due to landscape modification for infrastructure, deforestation and farmland on steep slope directly influence the hydrological conditions which affect the slope failure mechanism (Wang et al., 2020b). The landuse map of study area was prepared by analysing Landsat 8 OLI images through supervised image classification technique in ArcGIS environment. The output of landuse types were categorized into 5 classes i.e., barren land, forest, grass land, settlements and water body (Fig. 4f).

3.2.3.3 Precipitation

Various studies suggest that extreme events of precipitation affect the slope stability by triggering new landslides or reactivation of old landslide. Thus, the relationship of landslide occurrence with precipitation had been adopted as a critical LCF in predicting the landslide hazard around the globe (Huang et al., 2020). The effect of climate variations in study area during intense precipitation of monsoon season and winter westerlies increases the infiltration and erosion rate along vulnerable slopes which influence the probability of rapid slope failure (Rahman et al., 2014). In this paper, mean annual precipitation data of study area (2000-2020) were interpolated through inverse distance weighted (IDW) function in ArcGIS to generate the precipitation map. This map was further classified into four classes: <1000 mm, 1000-1200 mm, 1200-1400 mm and >1400 mm (Fig. 4m).

3.2.4 Anthropogenic factors

The anthropogenic activities such as construction of buildings, roads, terracing, stone quarries, etc., disturb the equilibrium state of slope material and increase the vulnerability of landslide hazard along steep slopes (Regmi et al., 2014). In this research, the proximity to roads was selected as the main anthropogenic factor influencing slope failures in the study area.

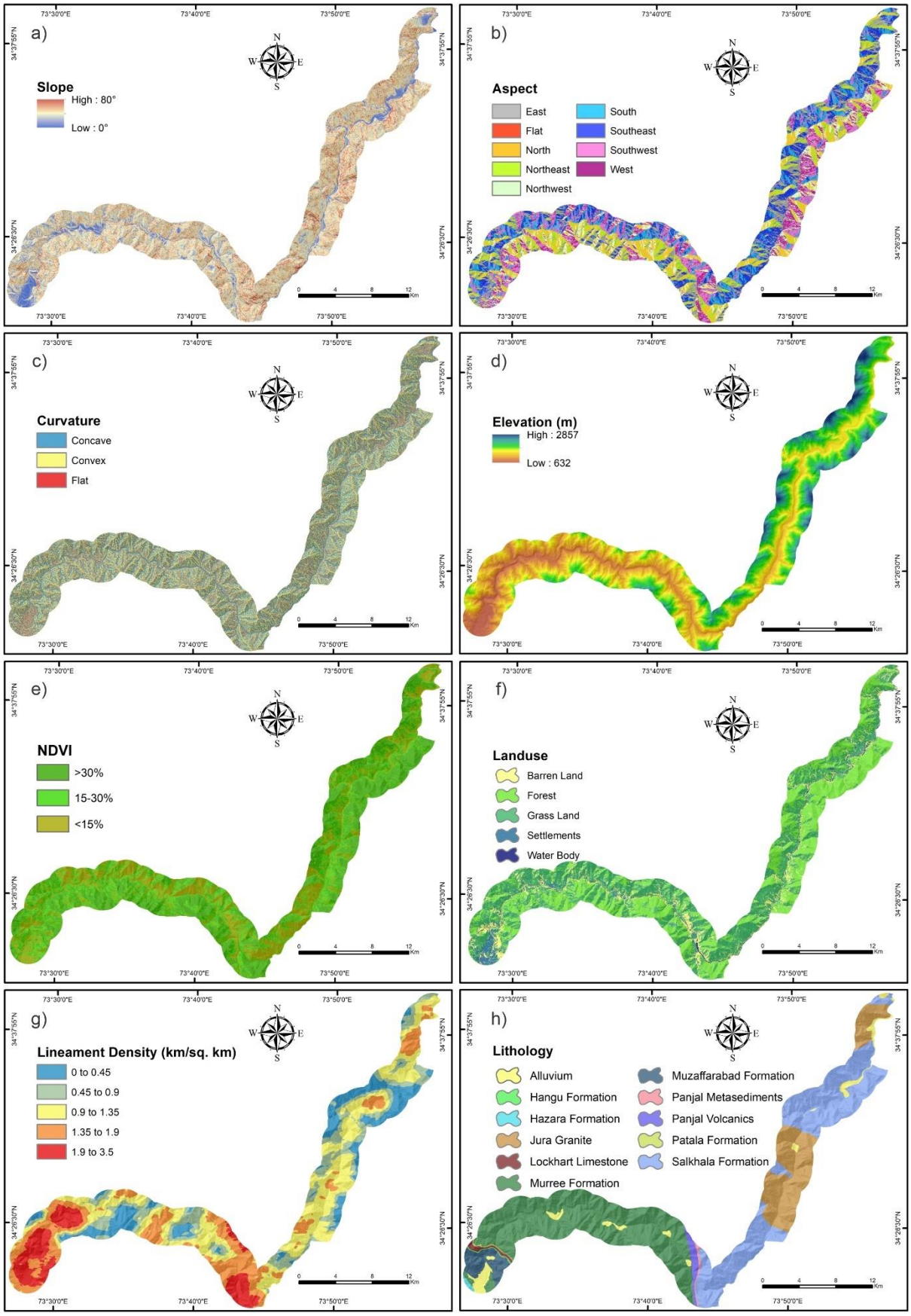

Fig. 4 Landslide causative factors utilized for landslide susceptibility prediction of study area

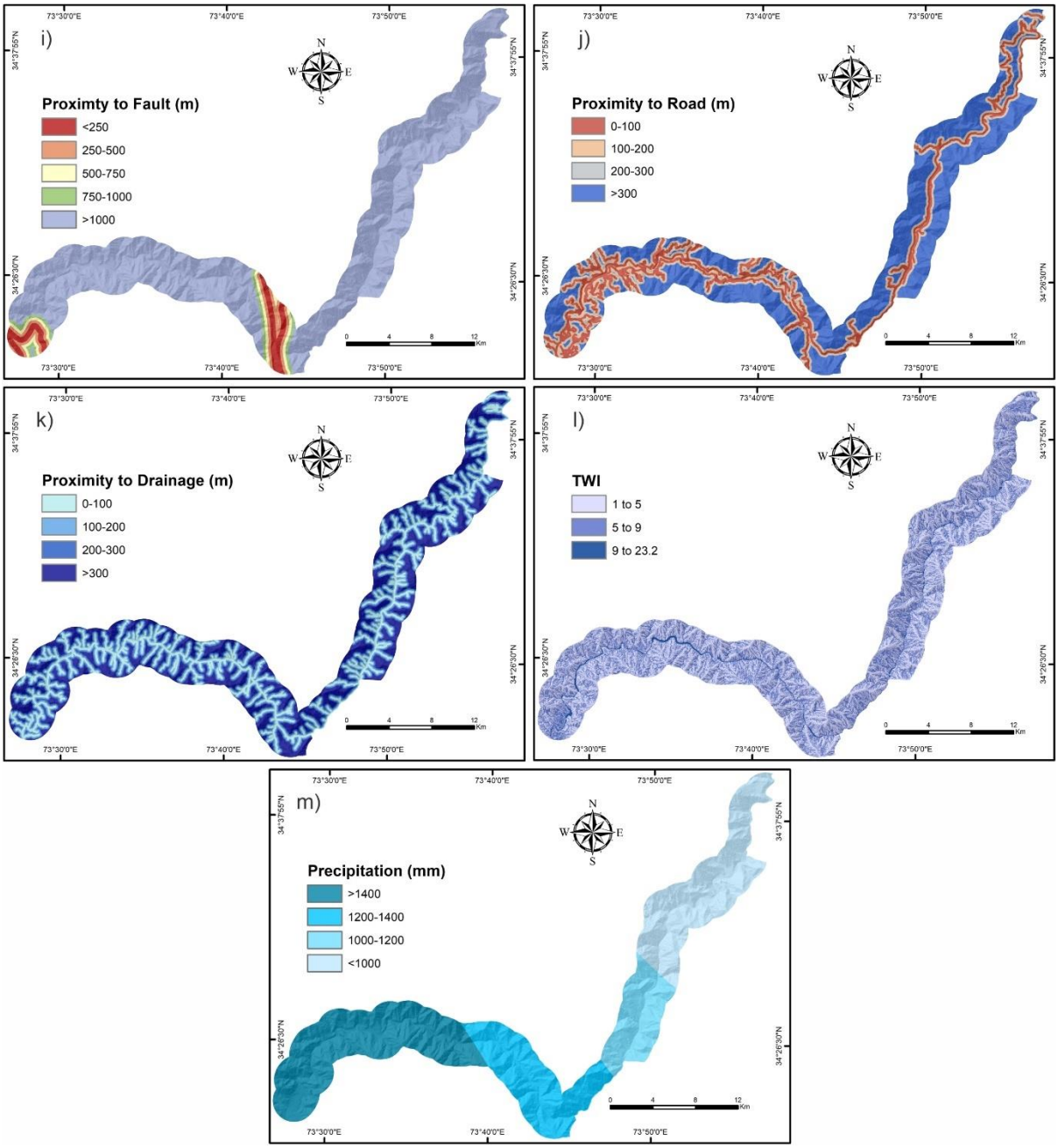

Fig. 4 (continued)

3.2.4.1 Proximity to roads

Roadcut being a site of anthropogenically disturbed slopes, can act as a catalyst for landslide occurrence in a mountainous region (Ayalew and Yamagishi, 2005). The excavation and obsolete blasting techniques for construction of roads alter the natural stability of slope and accelerate the probability of landslide incidences (Das et al., 2010). In present study, the landslide failure zones due to slope cutting were observed at a distance of approximately 100-

200m to major road network. Therefore, we have generated proximity to road map comprising of four buffer zones: 0-100m, 100-200m. 200-300 and > 300m (Fig. 4j).

4 Methodology

The methodological approach adopted for comparative assessment of different landslide prediction models is illustrated in figure 5 and categorically explained below.

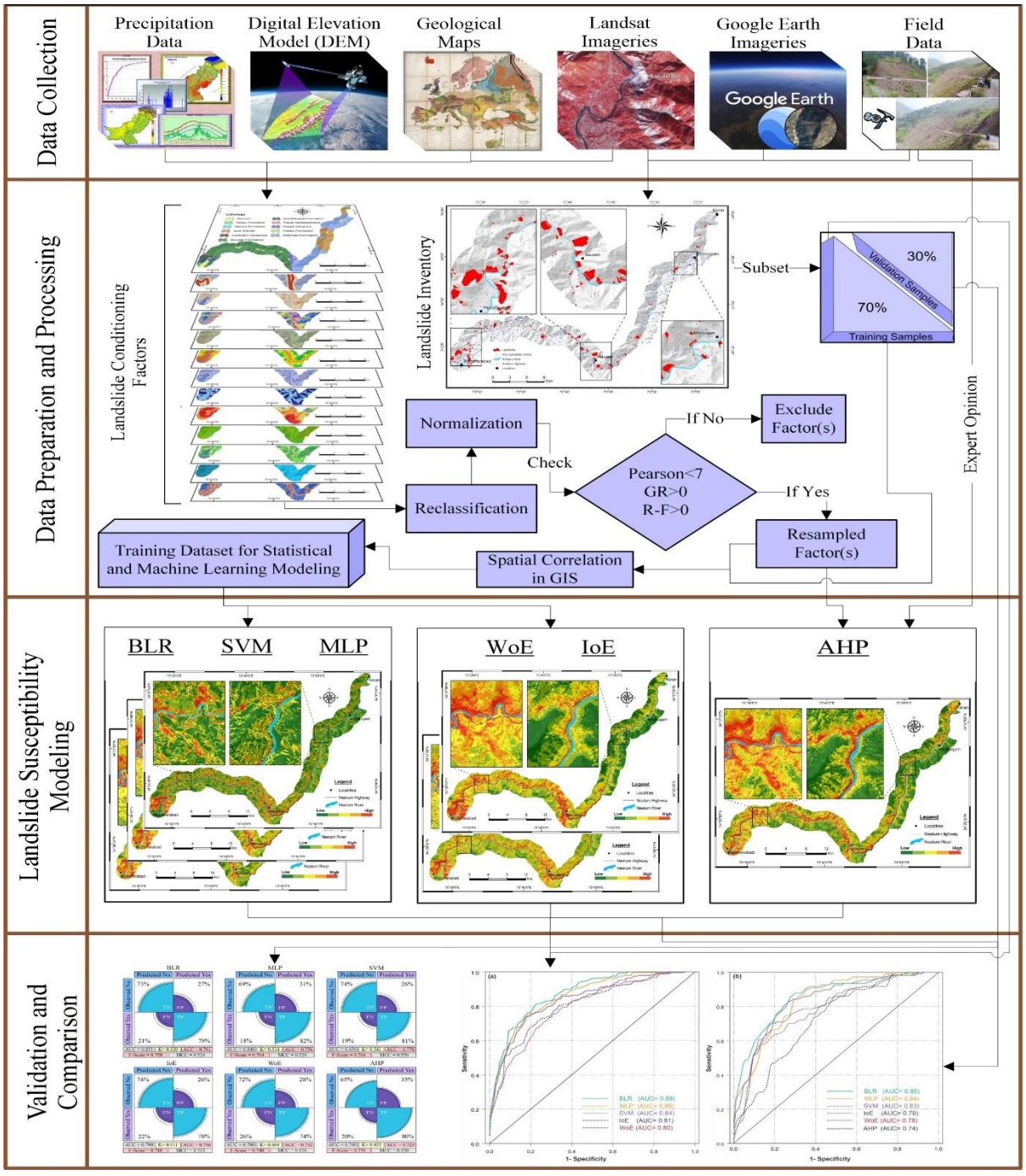

Fig. 5 Flow chart of applied methodology

4.1 Landslide predictors analysis

In this research, factor analysis of LCFs were adopted to identify the relevant predictors by measuring covariation among multiple variables and assessing the relative importance of causative factors in landslide occurrence.

4.1.1 Collinearity analysis

In landslide studies, collinearity among predictors may occur due to linear association of two or more independent variables which reduces the predictive performance of susceptibility models. Therefore, it is important to diagnose and eliminate the collinearity problem between input variables before data mining (Merghadi et al., 2020). To address this issue, variance inflation factor (VIF) and Pearson correlation were performed as statistical tests in this research.

VIF helps to quantify multicollinearity arise due to linear relationship between one variable and another variable in a model (Di Napoli et al., 2020). Mathematically, it is computed as:

$$VIF = (1 - R_i^2)^{-1} \quad (3)$$

here, R_i^2 is the coefficient of determination calculated by regression analysis of i^{th} predictors to other predictors. If VIF = 1, it indicates no correlation between one LCF to the remaining ones, whereas VIF >5 or 10 reflects higher multicollinearity (James et al., 2013).

On the other hand, the Pearson correlation is a statistical test that computes the extent of linear relationship between two predictors and their association with each other (Dou et al., 2020). The resultant linear association expresses how strong or weak relation exists between them. The coefficient of Pearson correlation (r) between two LCFs “a” and “b” is calculated as:

$$r_{ab} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{a_i - \bar{a}}{\sigma_a}\right)\left(\frac{b_i - \bar{b}}{\sigma_b}\right) \quad (4)$$

here, n refers to the size of data points, a_i and b_i denote the data points of a and b indexed with i, whereas $\bar{a}, \bar{b}$ and σ_a, σ_b are the means of observations and standard deviations of data points of variables a and b. The outcomes of the "r" range between -1 (negative correlation) and +1 (positive correlation), while zero value indicate no correlation among two variables. However, according to Nettleton (2014) the coefficient values of factors >0.7 represent high collinearity which can affect the model performance and should be eliminated before data modeling.

4.1.2 Factors contribution analysis

The landslides and their contributing factors vary from area to area depending on local geo-environmental conditions (Hong et al., 2020). Therefore, evaluating the role of different LCFs in terms of their importance for landslide occurrence will help to enhance the predictive capability of susceptibility models. For this purpose, the gain ratio and relief-F methods were applied in this study to estimate the relative importance of each conditioning factor in triggering of landslides.

Gain Ratio is well-known feature selection technique used to explore most important attributes in a dataset which render maximum information about a class. In landslide related studies, the contribution of each LCF is derived by finding the average merit (AM) values against corresponding factors. The highest AM scores represent the most relevant LCFs and vice versa, however a factor is regarded as irrelevant, if the AM values are ≤ 0 (Fallah-Zazuli et al., 2019). The gain ratio (GR) for a given LCF is computed as:

$$GR(A) = \frac{IG(A)}{E(A)} \quad (5)$$

here, IG (A) and E(A) are the information gain and entropy values with respect to an attribute "A". The function $E(A) = \sum_i -P(s_i)log_2P(s_i)$; where $P(s_i)$ indicates probability score of i^{th}

class by contributing to overall scores of attribute "A" classes. The IG function of a given attribute "A" with respect to class "X" is calculated in Eq. (6).

$$Info\ Gain(A) = H(X) - H_A(X) \tag{6}$$

where, H(X) represents the class "X" entropy and $H_A(X)$ is an average entropy of attribute "A" with respect to class "X" and are computed using Eq. (7) and Eq. (8), respectively:

$$H(X) = \sum_j -P(s_j) log_2 P(s_j) \tag{7}$$

$$H_A(X) = \sum_{j=1}^{m} P_j\ H(X) \tag{8}$$

in Eq. (7), $P(s_j)$ is probability score of j^{th} class corresponding to overall scores. Whereas in Eq. (8), m represents total number of class partitions, P_j is the probability score of j^{th} partition (Quinlan, 1993).

Relief-F is primarily used as data pre-processing technique that computes the quality and relevance of conditioning factor in a multiclass dataset based on correlation between predictors and their classes (Urbanowicz et al., 2018). Firstly, the relief-F randomly select an instance of a specific attribute, then look for the k nearest neighbours of same and different classes of selected instance termed as hits and misses, respectively. Finally, the rank of each predictor is evaluated by averaging the probability weights of all hits and misses (Kononenko, 1994). The computed scores of attributes in a given data are generally range between -1 (least significant) and +1 (most significant) (Kira and Rendell, 1992), however in landslide studies, if the factor contribution is equal to 0 or < 0, then it is considered as redundant factor and must be ignored in susceptibility modeling (Ali et al., 2021).

4.2 Landslide susceptibility models

In this paper, we have selected six popular models comprising of AHP, IoE, WoE, ANN, SVM and LR for predicting the landslide susceptibility of study area. The working mechanisms of these models are briefly described in the following subsections.

4.2.1 Heuristic model

AHP is the only heuristic approach we employed for the production of landslide susceptibility map. It is a well-known semi-qualitative method that involves expert based judgements to compute the criteria weights of different contributing variables through relative pair-wise comparison (Franek and Kresta, 2014; Saaty, 1990). The method has been successfully applied in various studies related to susceptibility mapping for landslides hazard assessments (Bera et al., 2019; Huang et al., 2020; Park et al., 2013; Shahabi et al., 2014; Yalcin et al., 2011). In our case, the practical framework adopted for computing weights of factor and their classes through AHP comprising of four stages i) hierarchic arrangement of causative factors, ii) pairwise comparative judgement (expert opinion) based on relative importance of factors using rating scale (1 to 9 and its reciprocal), iii) weight estimation of each parameter and its classes by computing the principal eigenvector, and iv) consistency checking to eliminate redundant judgments through consistency ratio (CR) which should be < 0.10 (Saaty, 1990)

The CR was calculated as the ratio of consistency index (CI) and random inconsistency index (RI). Here CI is defined as:

$$CI = {(\lambda_{\max} - n)}\Big/{(n-1)} \tag{9}$$

where, $\lambda_{\max}$ and n indicates the largest eigenvalue and total number of factors/ classes in a given matrix, respectively. Whereas, RI values were adopted from simulated average CI values of pairwise matrix specified by Saaty (1990) and Franek and Kresta (2014). Finally, the

landslide susceptibility index (LSI) was produced in accordance with weighted summation procedure using following equation:

$$LSI_{AHP} = \sum_{k=1}^{n} W_k \, w_{jk} \tag{10}$$

here, W_k denotes the weight value of LCF k; w_{jk} is jth class weight of factor k and n is total number of conditioning factors.

4.2.2 Statistical models

The statistical models can assist in predicting the susceptibility of landslide through various statistical operations by interpreting a relationship of landslides distribution with their causative variables in a given area (Huang et al., 2020). To comprehend the statistical approach for landslide susceptibility assessment in the study area, we utilized WoE and IoE models.

4.2.2.1 Weight of evidence

The WoE is well established and reliable geo-statistical technique evolved from information and Bayesian theory and is considered as an effective tool for assessment of landslide hazard (Saha and Saha, 2020). The core concept of WoE in landslide susceptibility modeling is to predict likelihood of landslide occurrence under same conditions by establishing a relationship between dependent (LCFs) and independent variables (presence and absence of landslide events (Ding et al., 2017). To achieve this target, firstly we calculated positive (W^+) and negative (W^-) weight scores for each class of LCFs using following equations:

$$W^+ = In \, \frac{P\{CF|LS\}}{P\{CF|\overline{LS}\}} \tag{11}$$

$$W^- = In \, \frac{P\{\overline{CF}|LS\}}{P\{\overline{CF}|\overline{LS}\}} \tag{12}$$

here, P is the probability; CF and $\overline{CF}$ indicates the presences and absence of respective causative factor; LS is the presence of landslide events and $\overline{LS}$ is the absence of landslide

events. Afterwards, the weight scores of W^+ were subtracted from W^- to find the weight contrast (W^c) of each class of LCF, which represents the spatial relationship of landslide events with classes of different causative factors. Finally, the LSI was calculated by the summation procedure using Eq. (13).

$$LSI_{WoE} = \sum_{i=1}^{k} {W^c}_{ij} \tag{13}$$

where, ${W^c}_{ij}$ is weight contrast score for j^{th} class of factor i and k representing total LCFs (Regmi et al., 2014).

4.2.2.2 Index of entropy

The index of entropy is bivariate statistical approach, successfully utilized by various investigators in the assessment of landslide susceptibility by measuring the extent of various causative factors (Wang et al., 2016). In landslide susceptibility modelling, IoE helps to estimate the weight values of each LCF influencing the occurrence of landslide using following equations:

$$P_{ij} = \frac{L_p}{D_p} \tag{14}$$

$$\left(P_{ij}\right) = \frac{P_{ij}}{\sum_{j=1}^{S_j} P_{ij}} \tag{15}$$

here, L_p is landslide percentage and D_p represents domain percentage; $\left(P_{ij}\right)$ denotes probability density; S_j indicates the number of classes. Whereas, H_j and H_{jmax} are the entropy values and they are expressed as:

$$H_j = -\sum_{j=1}^{S_j} (P_{ij}) \log_2\left(P_{ij}\right), j = 1, \ldots, n \tag{16}$$

$$H_{jmax} = \log_2 S_j \tag{17}$$

$$I_j = \frac{H_{jmax} - H_j}{H_{jmax}}, \qquad I = (0,1), \qquad j = 1,\ldots,n \tag{18}$$

$$W_j = I_j \times P_{ij} \tag{19}$$

Here in Eq. (18), I_j is the information coefficient, whereas W_j is the resultant weight values of each contributing factor in Eq. (19) (Hong et al., 2017; Liu and Duan, 2018). Finally, the LSI_{IoE} using IoE model, was prepared by following equation:

$$LSI_{IoE} = \sum_{i=1}^{n} \frac{Z}{m_i} \times C \times W_j \tag{20}$$

where, i is total number of LCFs used in mapping; the greatest number of classes denoted by Z; m_i represent the number of classes within a specific LCF; C is the secondary classified value of each class, whereas the W_j was obtained from Eq. (19).

4.2.3 Machine learning models

The Machine learning methods are well-understood statistical algorithms, efficiently explored in multiple disciplines to solve the complexity of big-data for making accurate predictions in contrast to traditional modeling techniques (Merghadi et al., 2020). In recent years, ML techniques have been extensively used to predict the susceptibility of landslides in various geo-environmental conditions with a significant level of predictive accuracy. In present case study, three popular ML approaches i.e. MLP, SVM and BLR were adopted to assess the landslide susceptible zones.

4.2.3.1 Multi-layer perceptron

The MLP is well known class of feed-forward neural network that uses a supervised learning algorithm to identify complex non-linear associations between predicting variables and response variables (Taud and Mas, 2018). The typical structure of MLP model consist of input/ visible layer, hidden layer and output layer interconnected with each other through a set

of neurons. The neurons present in theses layer are connected by corresponding weights which are trained to generate a desired output using activation function (Murtagh, 1991). In this study, we have implemented a single hidden layer MLP neural network to recognize the likelihood of landslide occurrences with respect to input LCFs. The schematic approach of this hierarchical structure is illustrated in figure 6. The learning parameters (epochs = 821, learning rate = 0.3) and activation function (sigmoidal) were calibrated according to (Tien Bui et al., 2016).

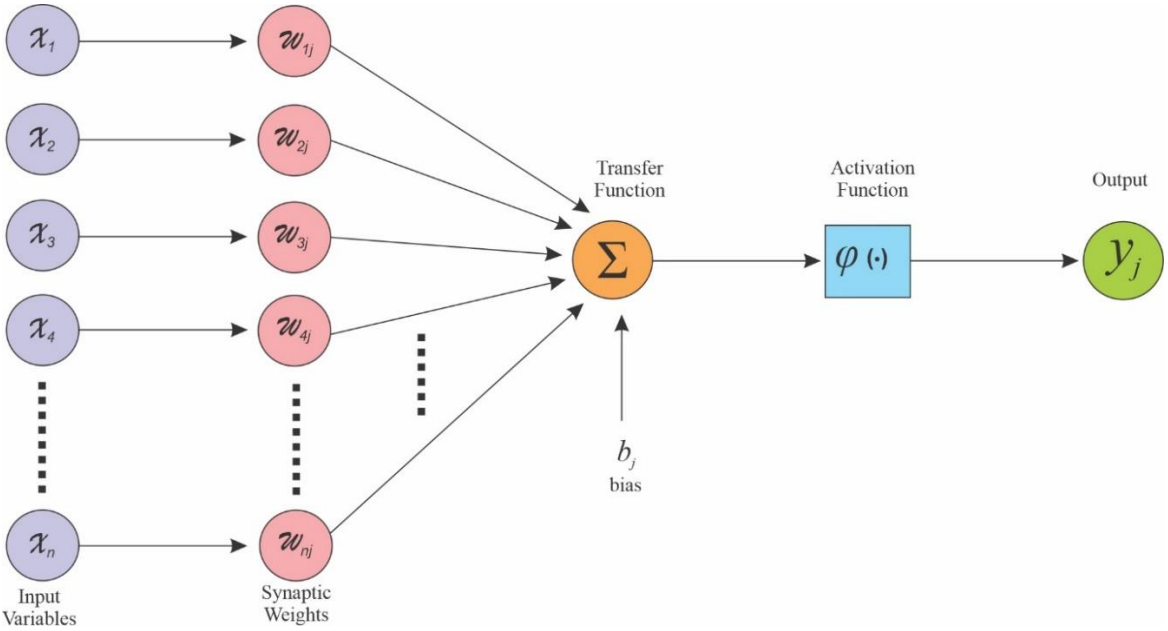

Fig. 6 Architecture of MLP model used in this study

4.2.3.2 Support vector machine

SVM is widely accepted as a robust predictive model in machine learning algorithms primarily used to solve classification problems. The core concept of SVM is to maximize the margin between classes and to segregate data points into different classes by establishing a suitable decision boundary or optimal hyperplane (Wu et al., 2008). In studies related to landslide susceptibility, it has been observed that the predictive performance of SVM model is better than other conventional techniques, however, the accuracy is associated with applied kernel function and other tuning parameters (Zhang et al., 2018). The radial base function (RBF) is proven to be an effective kernel function in LSP; therefore, RBF kernel was used in this study and can be expressed as:

$$RBF: K(x, y) = exp\{-\gamma \|x - y\|^2\} \quad (21)$$

here, K is kernel function; x and y are input vectors of landslide training data and γ is kernel parameter whose values ranges between 0 and 1. The optimum values of tuning or hyper-parameters such as C-regularization, kernel coefficient (γ) and loss function (ε) were adopted from (Merghadi et al., 2020). Afterwards, these parameters were employed in SVM algorithm using training dataset of normalized LCFs to model the susceptibility of landslides in this study.

4.2.3.3 Binary logistic regression

BLR is considered as one of most appropriate predictive ML algorithm used for classification of two-class dataset. It computes the probability of any event by establishing a relation between one target binary-variable and several predictors or independent variables (Lever et al., 2016). In this study, we used the presence (1) or absence (0) of landslides as target or dependent variable (y) and the LCFS as independent variable to map the probability (P) of landslide occurrence under the given conditions (Huang et al., 2020). Mathematically, the BLR model can be formulated as:

$$P_{BLR} = \frac{1}{(1 + e^{-y})} \quad (22)$$

$$y = \alpha_o + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n \quad (23)$$

where, α_o is the intercept of model; α_1, α_2, … α_n are the regression coefficients and x_1, x_2, ... x_n are the independent variables (LCFs).

4.3 Model validation and comparison

The evaluation of landslide prediction models is a core part of landslide susceptibility mapping for achieving an effective model performance and to ensure the scientific significance of models used. The validation of these models is commonly assessed through several evaluation metrics like area under receiver operating characteristics (AU-ROC) curve, cross-validation, confusion matrix, Root Mean Squared Error etc (Frattini et al., 2010). In this study,

we use AU-ROC curve, Matthew's correlation coefficient (MCC), Cohen's Kappa (K) coefficient, accuracy (ACC), F-score and confusion matrix fourfold-plot to evaluate model's accuracy and to discriminate among resultant models based on their performance scores. The AU-ROC curve was constructed by plotting sensitivity on horizontal axis and 1-specificity on vertical axis for both training and validation data, whereas the quantitative estimation of AU-ROC values was used to assess model quality; < 0.6 indicates poor results, > 0.8 represents very good results (Huang et al., 2020; Pourghasemi and Rahmati, 2018). The confusion matrix is kind of contingency table comprising of two rows and columns representing predicted values against actual values of both positive and negative landslide counts. This confusion matrix was also used to extract other important evaluation metrics such as ACC, F-score and MCC which can be useful for assessing robustness of applied models. The formulas for these evaluation indicators are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{24}$$

$$1 - Specificity = \frac{FP}{TN + FP} \tag{25}$$

$$AUROC = \frac{(\sum TP + \sum TN)}{P + N} \tag{26}$$

$$Accuracy = \frac{TP + TN}{TN + FN + TP + FP} \tag{27}$$

$$F\text{-}score = 2 \times \frac{Recall\ .Precision}{Recall\ + Precision} \tag{28}$$

$$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TP + FP).(TP + FN).(TN + FP).(TN + FN)}} \tag{29}$$

here P and N represents total number of landslides and non-landslide counts, respectively; TP (true positive) and TN (true negative) are correctly classified pixel counts; and FP (false positive) and FN (false negative) are incorrectly classified pixel counts (Achour and Pourghasemi, 2020; Chen et al., 2021) . The performance values for LSP models evaluated by

F-Score and accuracy varies from 0 to 1 (values close to 1 indicates as high predictive models and vice versa), whereas the maximum and minimum value of MCC range between +1 (best) and -1 (worst), respectively (Wang et al., 2021).

The Cohen's kappa coefficient measures the degree of agreement between observed (P_{obs}) and expected (P_{exp}) outcomes using following expression:

$$K = \frac{P_{obs} - P_{exp}}{1 - P_{exp}} \tag{30}$$

where, P_{obs} and P_{exp} are computed by following equations:

$$P_{obs} = {TP + TN}/{n} \tag{31}$$

$$\mathrm{P_{exp}} = \frac{(TP + FN)(TP + FP) + (FP + TN)(FN + TN)}{\sqrt{T}} \tag{32}$$

here, T is the total pixels used in training; n is the percentage of correctly classified pixels (Ali et al., 2021). The K values range between 0 and 1. The values close to 0 represent a random agreement, 0.21-0.40 indicates fair agreement, 0.41-0.60 imply a moderate agreement, 0.61-0.80 shows substantial agreement, whereas values >0.80 demonstrate a strong level of agreement (Cohen, 1960).

5 Results

5.1 Collinearity analysis

In case of landslide susceptibility modeling, the assessment of collinearity is important to diagnose unnecessary variables which may lead to decrease the predictive performance of resultant models. For this purpose, Pearson's correlation and VIF tests were performed to assess the degree of linear association among thirteen LCFs. The results of Pearson's corelation (all <0.7) indicates that all conditioning factors have low degree of association between them (Fig. 7).Similarly, the highest value of VIF is quite less than the threshold limit i.e., 5, (Wang

et al., 2020b) indicating no collinearity issue among all LCFs (Table 1). Therefore, the present analysis suggest that all thirteen predictors are suitable for susceptibility modelling in this study.

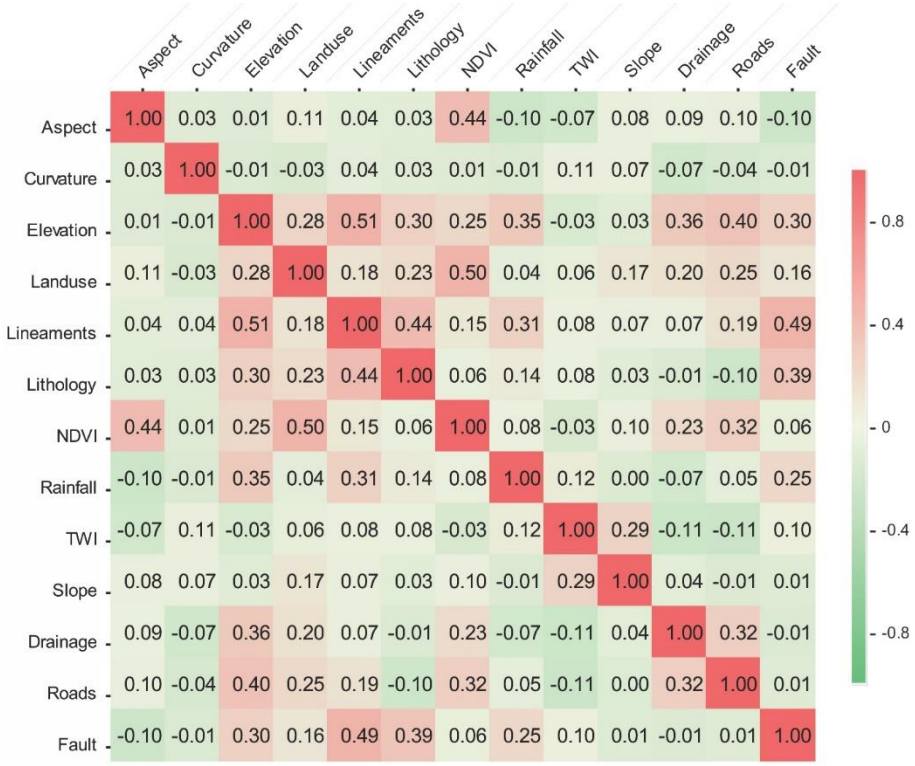

Fig. 7 Correlation chart of causative factors based on Pearson correlation matrix

Table 1 Result indicating collinearity diagnostic of LCFs using variance inflation factor

Sr. No.	Factor	Tolerance	VIF	Sr. No.	Factor	Tolerance	VIF
1.	Aspect	0.749	1.336	8.	Precipitation	0.781	1.281
2.	Curvature	0.975	1.026	9.	TWI	0.862	1.160
3.	Elevation	0.508	1.967	10.	Slope angle	0.878	1.139
4.	Landuse	0.657	1.523	11.	Drainage	0.768	1.302
5.	Lineaments	0.550	1.820	12.	Roads	0.699	1.431
6.	Lithology	0.684	1.463	13.	Fault	0.685	1.459
7.	NDVI	0.556	1.799				

5.2 Factor contribution analysis

Evaluating relative contribution of causative factors on occurrence of landslide is an important task because it helps to detect most significant predictors for efficient landslide susceptibility modeling. In present work, GR and Relief-F methods were implemented to perform this task. The calculated GR values for all thirteen LCFs shows that lithology (0.124),

landuse (0.110) and slope angle (0.084) are the dominant contributory factors, while lowest AM scores of TWI (0.009) and curvature (0.007) are within critical limit, therefore all factors were taken into consideration for landslide susceptibility prediction (Fig. 8a). Likewise, the obtained importance using Relief-F indicate that among thirteen LCFs slope angle (0.0877), landuse (0.0837), precipitation (0.0753) and lithology (0.0456) are the top most important factors, whereas TWI (0.0153), aspect (0.0130) and curvature (0.0055) are the least important factors in landslide probability modelling (Fig. 8b).

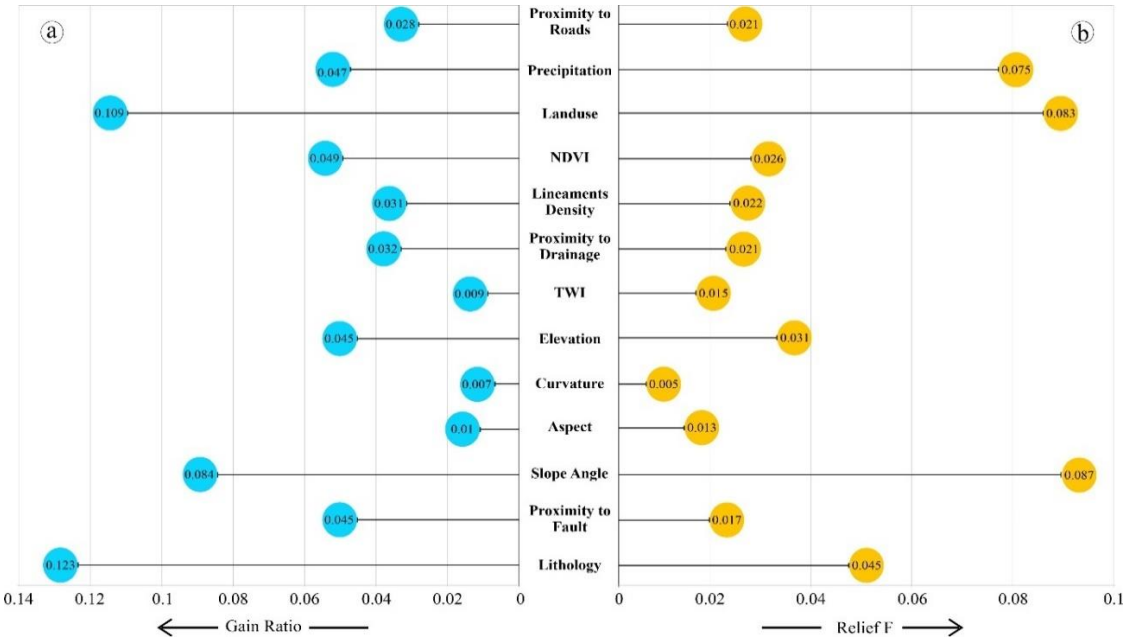

Fig. 8 Result of factor contribution analysis a) Gain Ratio b) Relief-F

5.3 Landslide susceptibility maps

The spatial prediction of landslide in the study area was explored by three different approaches i.e., heuristic (AHP model), statistical (WoE and IoE models) and machine learning (MLP, SVM and BLR models) to produce LSMs using thirteen landslide conditioning factors.

5.3.1 LSP using heuristic model

The class wise weights of each LCF in AHP analysis were computed through pairwise relative comparison matrix using knowledge driven approach based on extensive field studies. The result of this comparison matrix represents a reasonable consistency with values <0.10 for

all instances, therefore the calculated class weights of all thirteen parameters seem to be more realistic (Table 2). Similarly, the factor weights of all thirteen LCFs were evaluated by adopting the same technique. The CR for this analysis is 0.07 showing an acceptable level of consistency which implies that the calculated factor weights are reliable for predicting heuristic landslide susceptibility model. The final result of class weights and factor weights of 13 LCFs were used in Eq. (10) to produce LSI_{AHP} for the target area. The predicted LSI was reclassified into five classes using natural breaks classifier in GIS to produce LSM representing very low (22.2%), low (26.3%), moderate (26.7%), high (17.6%) and very high (7.2%) susceptible zones (Fig. 9, 11). On the other hand, quantitative analysis of actual landslides in each susceptibility class indicates that maximum percentage of landslide pixels lies in high to very high susceptible zones (Fig. 11).

Table 2 Weight values of causative factors obtained by pair-wise comparison matrix

Factor	Classes	Class weight	Factor weight
Lithology	Alluvium	0.026	0.23
	Murree Formation	0.278	
	Patala Formation	0.037	
	Lockhart Limestone	0.032	
	Hangu Formation	0.037	
	Panjal Metasediments	0.074	
	Panjal Volcanics	0.104	
	Jura Granite	0.066	
	Muzaffarabad Formation	0.212	
	Hazara Formation	0.003	
	Salkhala Formation	0.131	
	Consistency Ratio= 0.07		
Slope angle (°)	0-10	0.034	0.139
	10-20	0.049	
	20-30	0.075	
	30-40	0.104	
	40-50	0.189	
	50-60	0.332	
	> 60	0.217	
Proximity to fault (m)	0-250	0.445	0.168
	250-500	0.262	
	500-750	0.152	
	750-1000	0.089	
	>1000	0.052	
	Consistency Ratio= 0.006		
Lineament density	0-0.45	0.058	0.045
	0.45-0.9	0.09	
	0.9-1.35	0.156	
	1.35-1.9	0.256	
	1.9-3.5	0.44	
	Consistency Ratio= 0.01		
Aspect	Flat	0.024	0.019
	North	0.038	
	Northeast	0.06	
	East	0.087	
	Southeast	0.266	
	South	0.183	
	Southwest	0.216	
	West	0.095	
	Northwest	0.031	

Consistency Ratio= 0.032				Consistency Ratio= 0.074			
Curvature	Concave	0.592	0.016	TWI	1-5	0.26	0.029
	Flat	0.075			5-9	0.413	
	Convex	0.333			9-23.2	0.327	
Consistency Ratio= 0.015				CR= 0.06			
Elevation (m)	<1000	0.53	0.034	Landuse	Barren	0.525	0.055
	1000-1500	0.226			Forest	0.111	
	1500-2000	0.151			Grass land	0.233	
	2000-2500	0.058			Settlements	0.1	
	>2500	0.035			Water	0.031	
Consistency Ratio= 0.085				Consistency Ratio= 0.058			
Proximity to drainage (m)	0-100	0.37	0.077	Precipitation (mm)	<1000	0.09	0.068
	100-200	0.345			1000-1200	0.143	
	200-300	0.185			1200-1400	0.362	
	>300	0.1			>1400	0.406	
Consistency Ratio= 0.004				Consistency Ratio= 0.02			
Proximity to road (m)	0-100	0.467	0.09	NDVI	>15%	0.088	0.029
	100-200	0.277			15-30%	0.243	
	200-300	0.16			>30%	0.669	
	>300	0.095					
Consistency Ratio= 0.01				Consistency Ratio= 0.007			

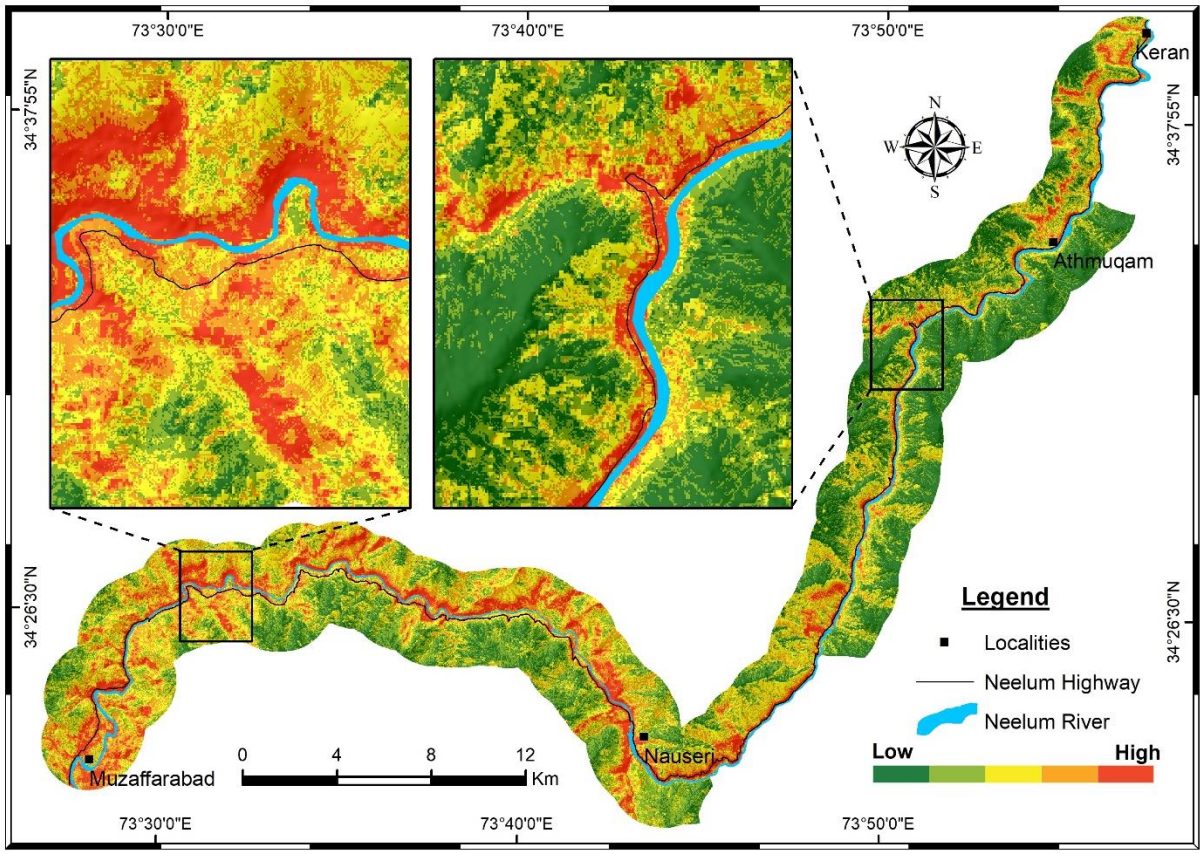

Fig. 9 landslide susceptibility map of study area produced by AHP model

5.3.2 LSP using statistical models

The WoE and IoE models were used as statistical approaches to compute class weights of each thematic layer by analysing a spatial relationship between landslide training dataset as dependent variable and all LCFs as independent variables (Table 3). The above analysis suggests that lithology, landuse, elevation, proximity to faults, lineament density and slope are most significant conditioning factors in triggering landslides, whereas the Muzaffarabad Formation, barren lands, moderate elevation, areas within 500m of faults, high lineament density and relatively higher slopes (40°-60°) are major factor classes influencing susceptibility of landslide occurrence in the study area.

The resultant LSIs for both statistical models were constructed by applying prediction scores in Eq. (13) and Eq. (20) using summation procedure in ArcGIS environment. These LSIs were than reclassified into five susceptible classes i.e., very low (VL), low (L), moderate (M), high (H) and very high (VH) using widely used Jenks classification scheme (Chen et al., 2021; Merghadi et al., 2020) to produce final LSMs of the study area (Fig. 10). The class wise area percentage distribution of landslide susceptible zones (LSZs) indicates that 75% area occupies VL to M zones, while the remaining 25% area falls in H to VH susceptible zones in WoE model. On the other hand, IoE model have slight changes in area percentage i.e., 77% in VL to M and 23% in H to VH susceptible zones (Fig. 11). Furthermore, the distribution of existing landslide pixels in each susceptibility classes depicts that most of landslide counts can be observed in H to VH susceptible zones of both statistical LSMs, representing a reliable prediction result (Fig. 11).

Table 3 Spatial relationship between causative factors and landslides derived from WoE and IoE models

Factors	**Classes**	**Domain Pixels**	**Landslide Pixels**	W^c	(P_{ij})	H_j	H_{jmax}	I_j	W_j
Lithology	Alluvium	75201	266	-2.504	0.065	2.626	3.459	0.24	0.363
	Murree Formation	721390	26400	-0.151	0.048				
	Patala Formation	3442	101	-0.323	0.044				

	Lockhart Limestone	11535	173	-1.013	0.022				
	Hangu Formation	3504	34	-1.451	0.014				
	Panjal Metasediments	8279	742	0.863	0.134				
	Panjal Volcanics	27893	4699	1.631	0.253				
	Jura Granite	329114	3605	-1.48	0.016				
	Muzaffarabad Formation	72495	16713	2.195	0.346				
	Hazara Formation	4376	10	-2.905	0				
	Salkhala Formation	583513	21056	-0.156	0.054				
Proximity to fault (m)	0-250	75172	7104	0.977	0.253	2.214	2.321	0.046	0.086
	250-500	58350	6607	1.181	0.303				
	500-750	43078	3263	0.696	0.203				
	750-1000	37667	2102	0.355	0.149				
	>1000	1626475	54723	-1.032	0.09				
Slope (°)	0-10	86272	818	-1.511	0.033	2.615	2.807	0.068	0.069
	10-20	180877	3616	-0.772	0.07				
	20-30	411071	12889	-0.318	0.11				
	30-40	686169	29565	0.122	0.151				
	40-50	390058	21270	0.428	0.191				
	50-60	73859	4897	0.559	0.232				
	> 60	12436	744	0.424	0.21				
Aspect	Flat	1202	3	-2.815	0.007	2.946	3.169	0.07	0.062
	North	235960	5626	-0.596	0.075				
	Northeast	218111	8211	-0.074	0.118				
	East	247603	9981	0.006	0.126				
	Southeast	297871	15948	0.373	0.168				
	South	225774	11637	0.305	0.162				
	Southwest	215073	11426	0.34	0.167				
	West	199845	8262	0.035	0.13				
	Northwest	199303	2705	-1.191	0.042				
Curvature	Concave	779455	34895	0.208	0.376	1.578	1.584	0.003	0.003
	Flat	287075	10842	-0.073	0.317				
	Convex	774212	28062	-0.175	0.305				
Elevation (m)	<1000	265155	26561	1.281	0.602	1.399	2.321	0.397	0.329
	1000-1500	803853	40051	0.443	0.299				
	1500-2000	584257	6360	-1.639	0.065				
	2000-2500	175944	818	-2.283	0.027				
	>2500	11533	9	-3.985	0.004				
TWI	1-5	1059972	38798	-0.211	0.295	1.579	1.584	0.003	0.003
	5-9	708662	31975	0.208	0.364				
	9-23.2	72108	3026	0.049	0.339				
Proximity to drainage	0-100	479148	26111	0.462	0.328	1.926	2	0.036	0.038
	100-200	417118	21224	0.335	0.306				
	200-300	343918	13281	-0.047	0.232				
	>300	600558	13183	-0.828	0.132				
Lineament density	0-0.45	271997	3966	-1.146	0.065	1.983	2.321	0.145	0.161
	0.45-0.9	355800	7050	-0.845	0.089				
	0.9-1.35	592984	18462	-0.367	0.139				
	1.35-1.9	402849	22570	0.474	0.251				
	1.9-3.5	215344	21751	1.222	0.453				
NDVI	>15%	731233	7465	-1.818	0.078	1.22	1.584	0.229	0.249
	15-30%	563948	21126	-0.1	0.286				
	>30%	545605	45208	1.386	0.634				
Landuse	Barren	207633	40905	2.479	0.647	1.559	2.321	0.328	0.498

	Forest	788307	7644	-1.924	0.031				
	Grass land	724298	19779	-0.593	0.089				
	Settlements	99340	5067	0.268	0.167				
	Water	21164	404	-0.77	0.062				
Rainfall (mm)	<1000	538418	10619	-0.929	0.126	1.872	2	0.063	0.062
	1000-1200	253749	6898	-0.454	0.173				
	1200-1400	405458	23964	0.557	0.377				
	>1400	641833	32317	0.391	0.322				
Proximity to road (m)	0-100	369875	28399	0.959	0.378	1.883	2	0.058	0.073
	100-200	247442	13547	0.387	0.269				
	200-300	186549	9240	0.249	0.244				
	>300	1036876	22613	-1.115	0.107				

5.3.3 LSP using machine learning models

In this paper, three standalone machine learning algorithms (MLP, SVM and BLR) were implemented in Weka 3.8.4 software to compute prediction scores of training dataset, which were than processed in GIS to build pixel-based LSIs of entire study area. The resultant LSI of each model was reclassified into very low, low, medium, high, and very high susceptible zones based on Jenks classifier to generate the map of landslide susceptible areas (Fig. 12). According to all LSMs, the higher susceptibility can be observed along Neelum highway, proximity to main faults and areas bordering Neelum river and its feeding tributes especially along the steep valley slopes in the central and western portions of study area. Overall, the relative area percentage, existing landslides distribution and their frequency ratio in each susceptible class of resultant LSMs are summarized in Figure 11. These result reveals that the area-based percentage of H to VH zones in all three ML models i.e., BLR (26.2%), SVM (25.2%) and MLP (25.3%) have almost similar proportion, whereas BLR (57.2%) gained higher proportion in VL to L susceptible classes as compared to SVM (47.8%) and MLP (46.4%) in the study area. Moreover, the frequency ratio analysis of past landslides in each susceptibility class reflects that maximum concentration of landslides were noticed in high to very high zones of all produced LSMs.

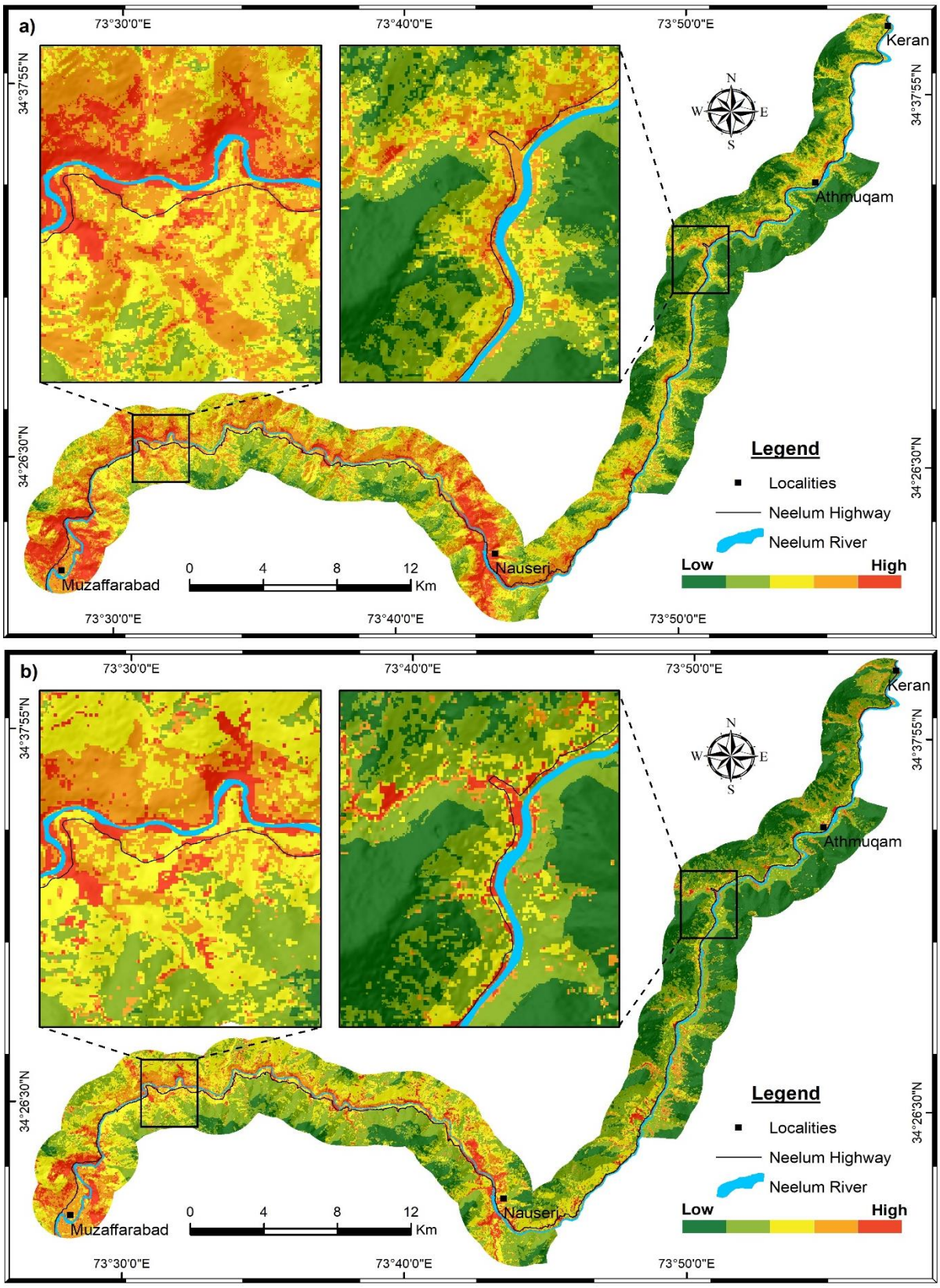

Fig. 10 Landslide susceptibility maps generated by statistical models a) WoE b) IoE

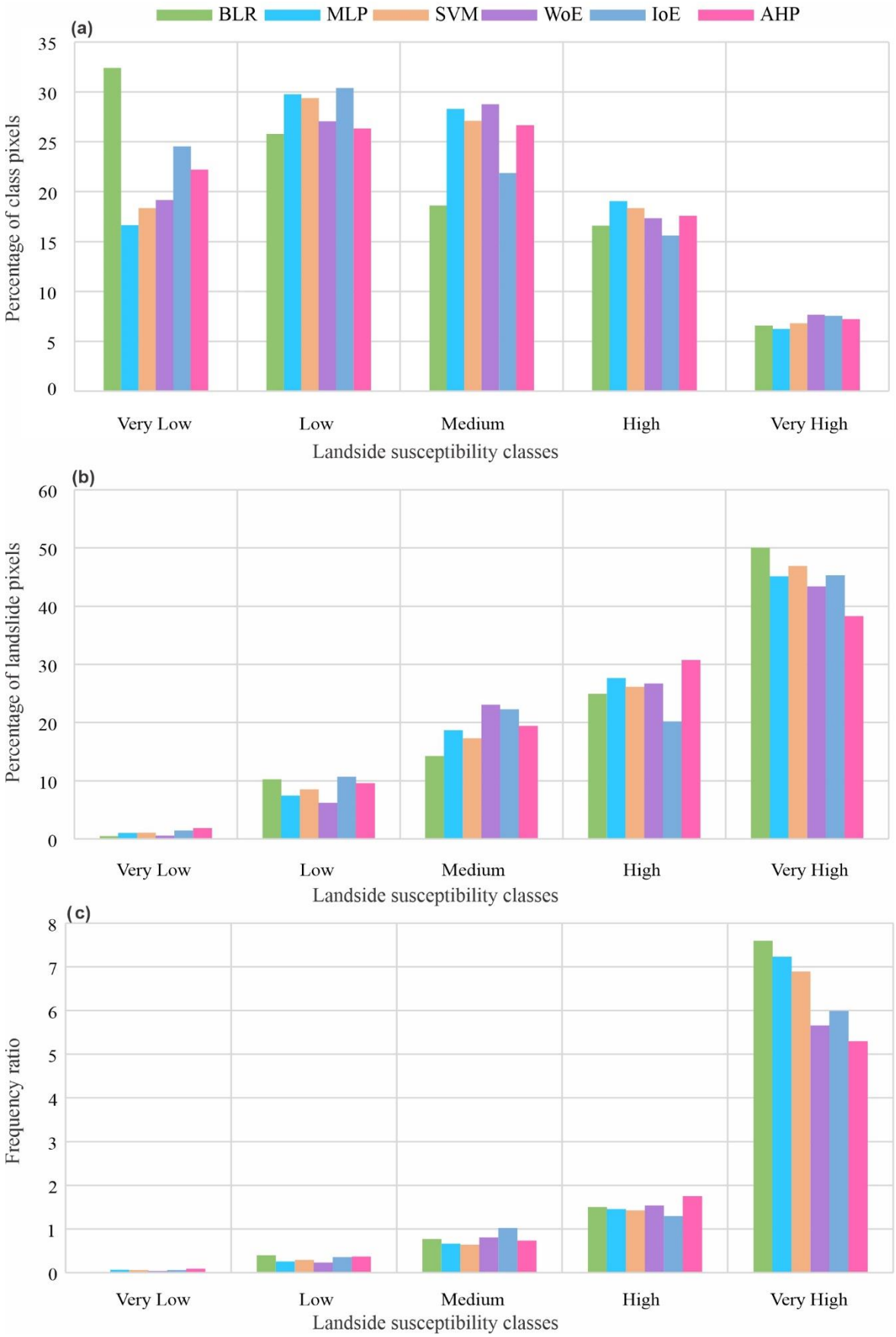

Fig. 11 Characteristics of five landslide susceptibility levels in all six LSMs based on a) area percentage; b) observed landslide percentage; c) landslide density

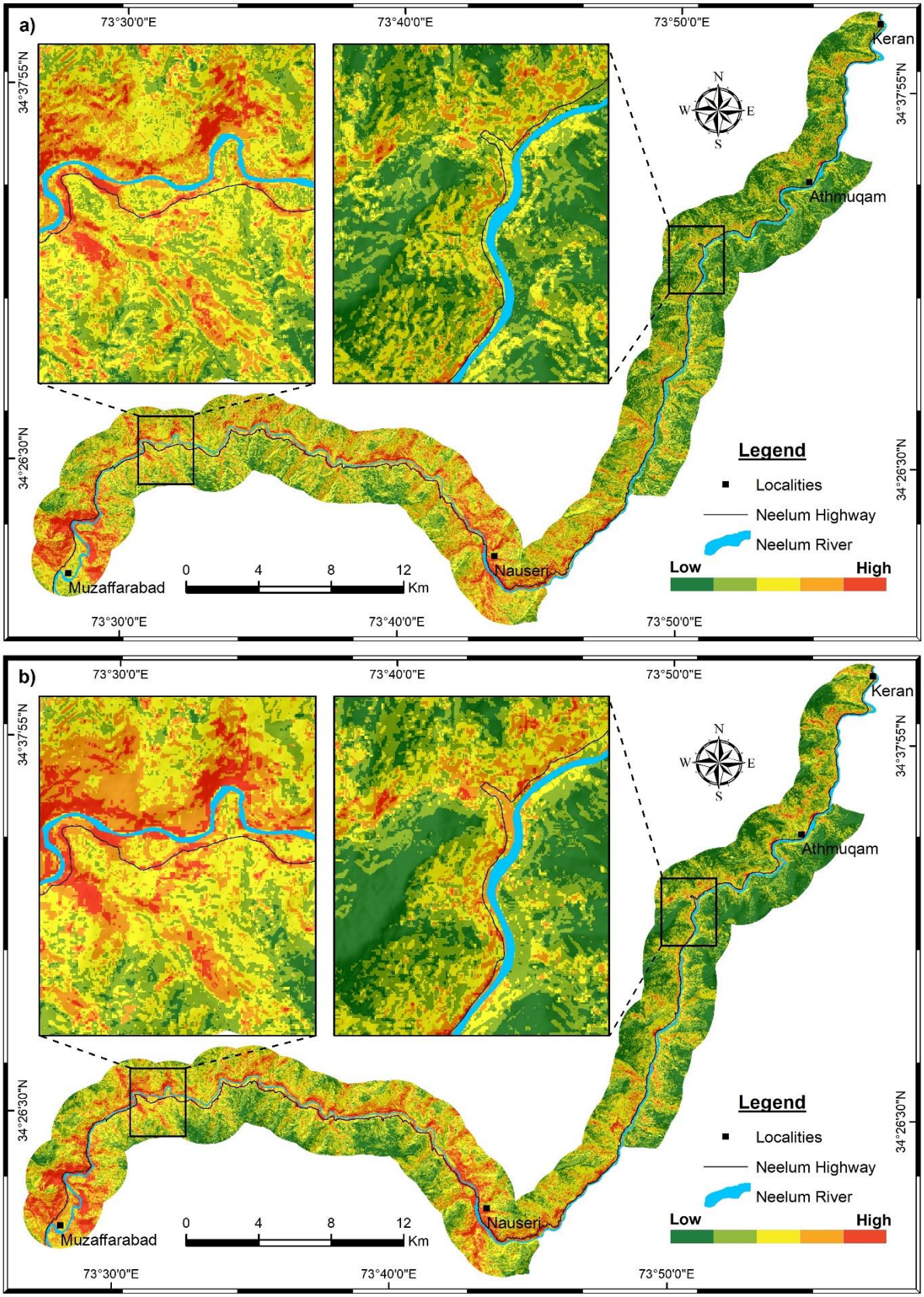

Fig. 12 Spatial distribution of landslide susceptibility constructed by using machine learning algorithms a) MLP; b) SVM; c) BLR

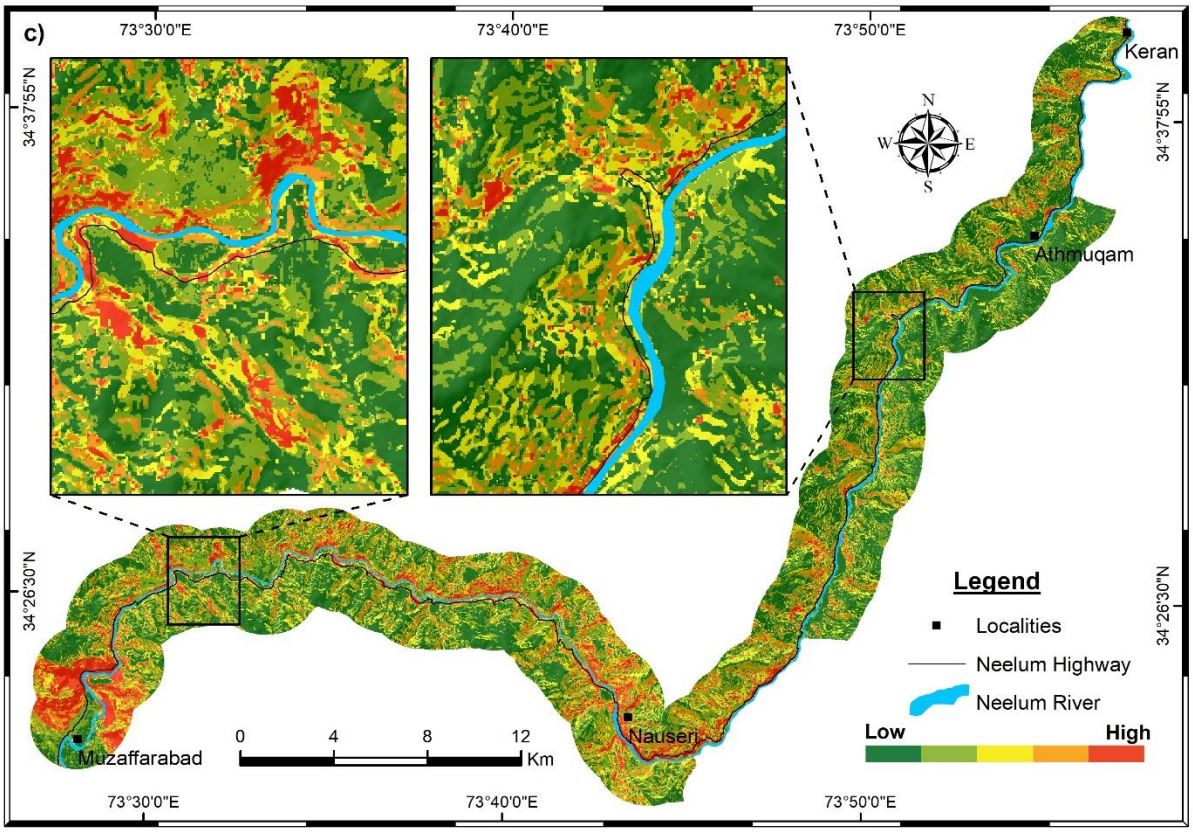

Fig. 12 (continued)

5.4 Model performance and comparison

The performance of each model employed in this study were assessed by examining the spatial concordance of actual data of landslide and non-landslide with the prediction results of each model using application of several validation methods i.e., AU-ROC, ACC, F-score, MCC, Kappa and confusion matrices (Fig. 13, 14). The training dataset was utilized to compute the performance of each model in terms of their success rate, whereas the predictive ability of derived LSP models was evaluated by using testing dataset of observed landslides and non-landslides.

According to the results, respective AU-ROC values of 0.881, 0.860, 0.842, 0.810 and 0.802 for BLR, MLP, SVM, IoE and WoE models showed that the success rates of LSP models generated by BLR, MPL and SVM were superior as compared to IoE, WoE and AHP (Fig. 13a). Likewise, the ACC and F-score results of BLR (0.797, 0.787), SVM (0.777, 0.764), MLP (0.773, 0.753), IoE (0.755, 0.739) and WoE (0.741, 0.745) models suggest that BLR model

had performed better as compared to others. An identical order was observed in the results yielded by MCC and K coefficient matrices which implies that BLR (0.609, 0.604) model had gained highest performance rate followed by SVM (0.561, 0.554), MLP (0.554, 0.547), IoE (0.520, 0.511) and WoE (0.491, 0.489) models (Table 4).

On the other hand, the prediction scores obtained through AU-ROC for all six models indicates a similar tendency with slight variations in results when compared with performance scores of training data. The BLR model achieved highest AU-ROC value of 0.85 followed by MLP (0.84), SVM (0.83), IoE (0.79), WoE (0.78) and AHP (0.74) models (Fig. 13b). Whereas, different pattern was observed in the results of ACC and F-score based on testing dataset i.e., SVM (0.770, 0.766), BLR (0.761, 0.764), MLP (0.756, 0.758), IoE (0.756, 0.748), WoE (0.732, 0.740) and AHP (0.723, 0.731). Similarly, the results of MCC and K coefficient matrices showed that predictive performance of all models were lies in good category and have an acceptable level of agreement. The highest and lowest values of these metrics range between range 0.554 to 0.450 (Table 4). Furthermore, the contingency plots of all six models illustrating percentages of correctly and incorrectly classified landslide and non-landslide picture elements validate the reliability of employed models in predicting spatial distribution of landslide in study area (Fig. 14). The overall performance scores comparison of all six models indicates that BLR, MLP and SVM (machine learning models) had significantly higher learning and prediction capabilities than statistical (IoE and WoE) and AHP models, respectively.

Table 4 Comparison of model performance using training and testing datasets

Model	Training Dataset					Testing Dataset				
	AUC	ACC	F-score	MCC	K	AUC	ACC	F-score	MCC	K
BLR	0.881	0.797	0.787	0.609	0.604	0.851	0.761	0.764	0.554	0.549
MLP	0.860	0.773	0.753	0.554	0.547	0.840	0.756	0.758	0.520	0.514
SVM	0.842	0.777	0.764	0.561	0.554	0.830	0.770	0.766	0.550	0.541
IoE	0.810	0.755	0.739	0.520	0.511	0.790	0.756	0.748	0.513	0.511
WoE	0.802	0.741	0.745	0.491	0.489	0.780	0.732	0.740	0.470	0.464
AHP	-	-	-	-	-	0.740	0.723	0.731	0.450	0.455

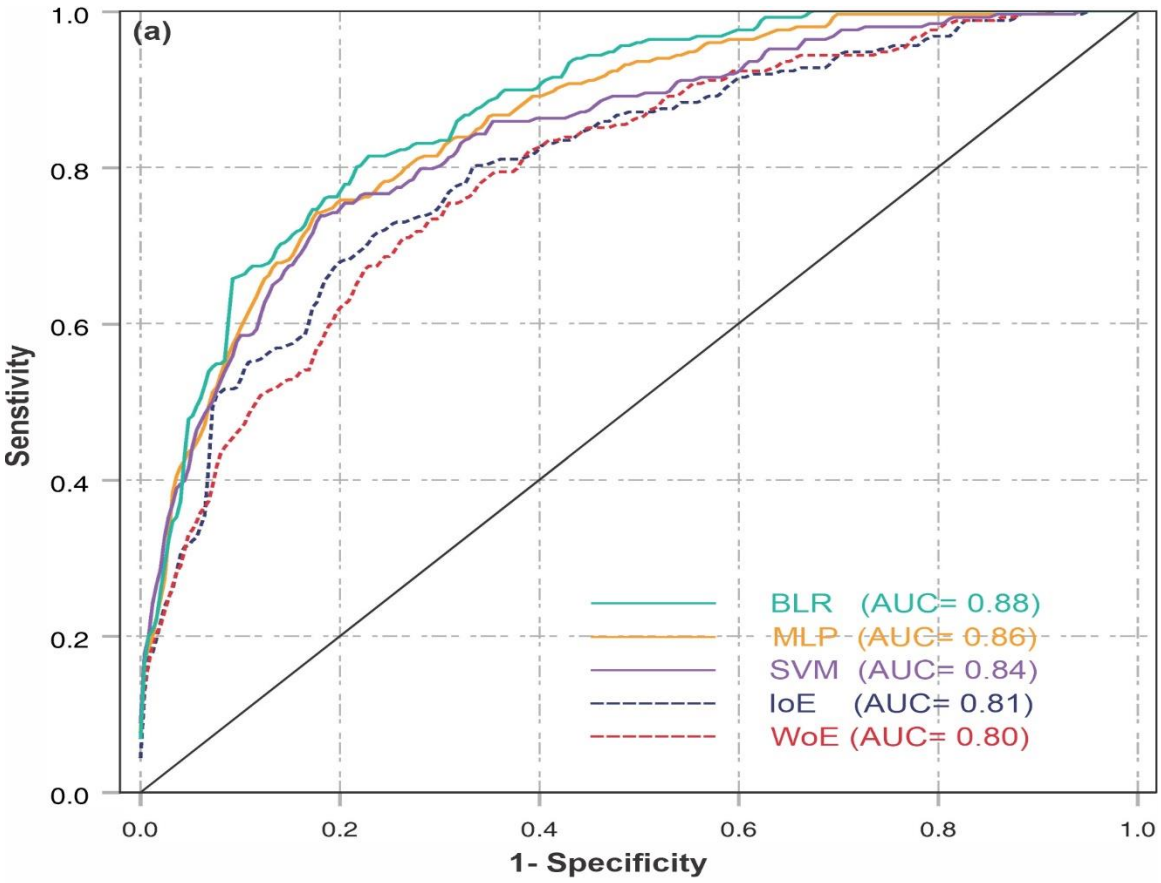

Fig. 13 ROC curves of all six models a) training dataset; b) validation dataset

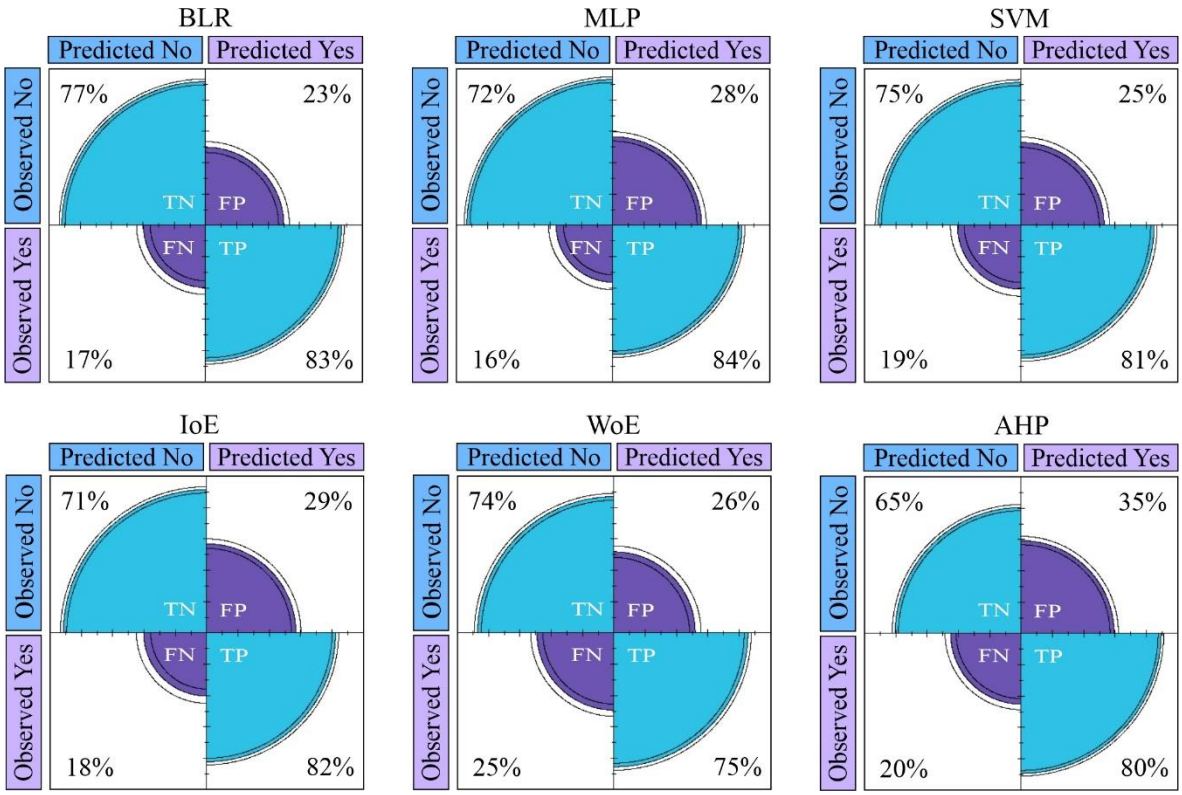

Fig. 14 Contingency plots summarizing percentage of TN, TP, FN and FP for employed models

6 Discussion

The spatial prediction of landslide susceptibility at various scales provides a foundation for effective landuse planning and hazard management in a complex mountainous landscape (Aditian et al., 2018; Erener and Düzgün, 2012). Owing to its unique topographic configuration, tectonic fragility and harsh climatic conditions, the Kashmir Himalaya is considered as a place sensitive for landslide disasters. Being a part of this region, the studied area has faced an enhanced rate of landsliding in recent decades due to mega earthquake of 2005 and rapid growth in socio-economic activates, which exacerbate the risk of landslide hazard (Rahman et al., 2014; Sana and Nath, 2017). However, the exploration of LSP models using current practices and their comparison with traditional techniques to evaluate an optimal model for landslide susceptibility assessment is still lacking in this region. To address this issue, we employed and compared widely used data-driven techniques such as machine learning (MLP, SVM and BLR), statistical (WoE and IoE) and heuristic (AHP) to generate LSMs of earthquake affected road-section, a part of western Himalaya in Kashmir.

In general, the LSP modelling is mainly dependent on selection of factors contributing landslides, as it has a substantial impact on predictive performance of models. The criteria for the selection of LCFs is still debatable because the variable that influence landslides may varies under different geo-environmental conditions. However, it is important to employ variable selection techniques in order to maximise the performance of LSP model by removing noisy factor before training phase (Merghadi et al., 2020; Reichenbach et al., 2018). To ensure this, Pearson correlation and VIF tests were performed which demonstrates that degree of liner association among all the thirteen LCFs lies within the threshold limits specified by Dou et al. (2020) (Fig. 7; Table 1). Additionally, the selected LCFs were analysed using GR and Relief-F methods to find the contribution of each factor in aiding to initiate landslides. Results revealed that the slope angle, lithology, landuse and precipitation were the most significant factors while curvature, TWI and aspect were the least important factors influencing landslides in study area (Fig. 8). Our findings are an agreement with the previous studies (Basharat et al., 2014a; Basharat et al., 2016; Kamp et al., 2008; Rahman et al., 2014), where these factors played a dominant role in occurrence of landslides. Similarly, the distribution of landslides and their causes inferred from field observations also complement our findings; for instance, fragile lithology on terrain slopes aggravated the phenomena of landsliding when it rains and the rapid and unplanned construction activities also triggered shallow landslides. On the basis of these analysis, 13 LCFs i.e., lithology, proximity to faults, slope angle, aspect, curvature, elevation, TWI, proximity to drainage, lineaments density, NDVI, landuse, precipitation and proximity to roads were selected in this article to predict landslides susceptibility.

Numerous studies were performed to generate regional scale LSMs and proposed suitable models for the assessment of susceptibilities. Recently a review highlighted that more than 500 research articles based on 70 different models were published in 2005-2016 (Pourghasemi et al., 2018), however it is pertinent to mention here that few studies (Ahmed et al., 2021; Basharat

et al., 2016; Kamp et al., 2008) were conducted to generate LSMs of 2005 earthquake affected areas of Kashmir Himalayas. Additionally, these investigations have a limited scope in respect of modelling techniques and lacking current practices with an integrated approach to demarcate the landslide susceptible zones. In order to generate the LSMs of area under investigation, six different data-driven models (AHP, BLR, IoE, MLP, SVM and WoE) were built by utilizing expert opinion and training dataset along with validation dataset. Thus, these models were employed to determine the LSMs of all the pixels, and then compared with each other in terms of prediction accuracies and distribution of landslide susceptibility levels in corresponding LSMs to analyse more accurate LSP model. Previously, Basharat et al. (2016) and Kamp et al. (2008) produced LSMs, which demonstrated the landslide potential zones along the major faults and/or associated with seismic events. Furthermore, Khan et al. (2013) documented landsliding on the basis of 5 years photographic record (2005-2010) in this region and concluded that landscape resumed to equilibrium in five years after earthquake, although the phenomena of landsliding is still ongoing which has a positive relationship with human activities. We support the findings of Khan et al. (2013) that earthquake triggered landslide became stable and/or exhausted in 2010, while in recent years a large number of new landslides occurred or reactivation of existing landslides is observed which could be result of intense rainfalls, erosion by Neelum river and most influentially by rapid and unplanned increase in slope cutting for socio-economic activities. Considering new landslides (2010-2020) in addition to the pre-existing landslides data along targeted road-section, our LSMs predicted the landslide potential zones with a high rate of predictive accuracy as compared to previous studies. The generated LSMs of all six models are well harmonized with one another, although the ratio of LSZs with respect to area percentage distribution is different in each model (Fig. 11). Generally, a uniform pattern of high and very high LSZs in all models observed along an earthquake affected road section of Neelum Highway, however it was noticed in field

investigations that landsliding is the most common phenomena with potential hazards along slope angle ranging from 40° to 60° and mainly observed in the fragile lithologies especially along faults, roads, and drainages. On the other side, the low and very low LSZs are mainly dispersed in the areas with slope angle less than 20°, far from roads and drainage networks, away from faults and less affected by anthropogenic activities. In general, the machine learning models i.e., BLR, MLP and SVM predicted well with respect to observed landslide percentages of 75.0%, 73.04% and 72.79% for high and very high LSZs in comparison with statistical models (WoE= 70.11%, IoE= 68.54%) and AHP model (69.06%) of area under study (Fig. 11).

Furthermore, different accuracy assessment methods i.e., AU-ROC, ACC, F-score, MCC, Kappa and confusion matrices on testing and training datasets were employed for each model to evaluate/validate the model prediction performance and their comparison. The model with a greater and/or lower success rate and prediction rate can validate and explain how well the model classified the area based on landslide inventory and how well the model predicted the landslide hazards (Pourghasemi and Rahmati, 2018). As a result of these analysis, it can be deduced from the model's comparison in section 5.4 that all the six models used in this study have shown a considerable predictive capability to produce LSMs, although the machine learning models in general have higher rate of predictive accuracy when compared with other models. In addition to this, the machine learning algorithms have many benefits in the model processing and relatively simple to interpret without bulk of background information (Achour and Pourghasemi, 2020; Wang et al., 2021). Whereas, the other data-driven techniques used in this study need several datasets to train and to test each model. Unfortunately, not a single machine learning technique model was used for landslide susceptibility mapping in the seismically active zone of studied region. Only Kamp et al. (2008) produced a regional level LSM of 2005 earthquake affected areas with a prediction accuracy of 67% by using AHP technique, whereas our findings revealed 74% accuracy of landslide prediction by AHP model

which has a considerably better degree-of-fit to the assessing data. Overall, BLR method outperformed the other models and exhibits a highest capability in terms of landslide prediction and LSZs distribution, followed by MLP, SVM, IoE, WoE and AHP model (Fig. 13; Table 4).

Being an important portion of the discussion section, the limitations of the techniques and dataset employed in this study can be summarised as; i) currently, there are no standards and code of practices established to map and evaluate the susceptible areas for landslide occurrence. Therefore, it is always a challenge in selection and adoption of an appropriate technique/ method for the evaluation of landslide susceptibility, which resulted in an uncertain condition on account of its reliability and credibility of outcomes obtained by the adopted method. ii) in this study, we were focused on earthquake affected road-section to generate LSMs and to compare their predictive performances. For this, landslide inventory was prepared by compiling field-data and remote sensing data. However, due to limitations of historical images, we were not able to exactly identify the landslides before 2005 Kashmir earthquake particularly of smaller size. iii) another limitation of this study is that it doesn't propose the causative mechanism of slope failure due to a complex relationship with triggering factors, thus it can only be employed as an initial step in assessment of landsliding phenomena. iv) the analysis based on AHP model is one of the limitations, as it strongly relies on opinion of geologist/expert and it refers to their subjectivity/ field knowledge in a specific area. v) an important problem can be seen that we didn't classify landslides by their type of movement in order to generate LSMs and considered each type of landslide occurred by following same processes, whereas the causative factors and triggering factors may vary for different mode of movement. vi) finally, the last and most important point is that the results obtained for the assumption of landslide hazard zones are only viable for similar geo-environmental conditions. Therefore, the predictions based on present geo-environmental conditions and/or the

mechanisms that controls the slope failure are not a guarantee for the future, if these conditions change.

To improve the presented methodology of landslide susceptibility zonation, above mentioned limitations (i) to (v) must be considered for developing new LSP models. Recently, novel machine learning and deep learning approaches as well as their ensembles explored in landslide susceptibility mapping with significant prediction accuracies. So, employing these techniques and comparing with other statistical and multicriteria decision making methods can help to find an optimal model for LSP in studied region of Kashmir Himalayas.

7 Conclusion

The landslide disasters in earthquake prone region of Kashmir Himalayas are mainly controlled by combined interaction of multiple geo-environmental and anthropogenic drivers which badly affected the human population in terms of infrastructure deterioration and socio-economic development. Hence, it is indispensable to assess the critical vulnerable zone of slope failures in this region for emergency management /risk analysis and minimizing oncoming landslide hazards. The present study was focused to analyse landslide susceptibility along a road-section of Neelum highway where natural slopes were severely disturbed by post-earthquake activities. We developed six different LSP models by applying heuristic (AHP), statistical (WoE and IoE) and machine learning (BLR, MLP and SVM) techniques based on site-specific causative factors with detailed landslide inventory and compared their prediction performance to evaluate robustness of employed models. Our analysis revealed that all the six models show promising results, however machine learning models exhibited remarkable LSP in comparison with statistical and heuristic models. The BLR reflected highest prediction accuracy in a seismogenic zone of Kashmir Himalaya followed by MLP, SVM, WoE, IoE models, while AHP was found to be least effective method in demarcation of landslide susceptible areas along targeted road-section. The performance scores derived through

statistical metrics provides a valuable indication that the employed models performed well in predicting spatial representation of landslide susceptibility, however, the machine learning portrays highest predictive performance, comparatively. The majority of predicted landslides distribution pattern was observed in western and central part of study area, particularly in the proximity of road/faults and deeply dissected valley slopes which replicates the actual situation of slope failures evident from field observation. The analysis of factor contribution in landsliding and field observations suggested that unstable and fragile lithology on steep slopes were dominant contributors in landsliding mainly controlled by climatic and anthropogenic factors. In conclusion, we presumed that the established methods have a reliable capability to predict landslide potential zones, which could assist the local government and decision makers in mitigation and prevention of landslide disasters along the highway road-corridors.

8 References

Achour, Y. and Pourghasemi, H.R., 2020. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? Geoscience Frontiers, 11(3): 871-883.

Aditian, A., Kubota, T. and Shinohara, Y., 2018. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology, 318: 101-111.

Ahmed, K.S., Basharat, M., Riaz, M.T., Sarfraz, Y. and Shahzad, A., 2021. Geotechnical investigation and landslide susceptibility assessment along the Neelum road: a case study from Lesser Himalayas, Pakistan. Arabian Journal of Geosciences, 14(11): 1019.

Aleotti, P. and Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the Environment, 58(1): 21-44.

Ali, S.A., Parvin, F., Vojteková, J., Costache, R., Linh, N.T.T., Pham, Q.B., Vojtek, M., Gigović, L., Ahmad, A. and Ghorbani, M.A., 2021. GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. Geoscience Frontiers, 12(2): 857-876.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology, 65(1): 15-31.

Baig, M.S., 2006. Active faulting and earthquake deformation in Hazara-Kashmir syntaxis, Azad Kashmir, northwest Himalaya, Pakistan, Extended Abstracts, International Conference on 8 October 2005 Earthquake in Pakistan: Its implications and hazard mitigation. Geol. Survey of Pakistan Islamabad, pp. 27-28.

Basharat, M., Rohn, J., Baig, M.S. and Khan, M.R., 2014a. Spatial distribution analysis of mass movements triggered by the 2005 Kashmir earthquake in the Northeast Himalayas of Pakistan. Geomorphology, 206: 203-214.
Basharat, M., Rohn, J., Baig, M.S., Khan, M.R. and Schleier, M., 2014b. Large scale mass movements triggered by the Kashmir earthquake 2005, Pakistan. Journal of Mountain Science, 11(1): 19-30.
Basharat, M., Shah, H.R. and Hameed, N., 2016. Landslide susceptibility mapping using GIS and weighted overlay method: a case study from NW Himalayas, Pakistan. Arabian Journal of Geosciences, 9(4): 292.
Bashir, F. and Hanif, U., 2018. Climate change extremes and rainfall variability over pakistan, Global Change Impact Study Centre – Pakistan.
Bera, S., Guru, B. and V, R., 2019. Evaluation of landslide susceptibility models: A comparative study on the part of Western Ghat Region, India. Remote Sensing Applications: Society and Environment, 13: 39-52.
Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrological Sciences Bulletin, 24(1): 43-69.
Bossart, P., Dietrich, D., Greco, A., Ottiger, R. and Ramsay, J.G., 1988. The tectonic structure of the Hazara-Kashmir Syntaxis, southern Himalayas, Pakistan. Tectonics, 7(2): 273-297.
Bui, D.T., Tsangaratos, P., Nguyen, V.-T., Liem, N.V. and Trinh, P.T., 2020. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. CATENA, 188: 104426.
Calkins, J.A., Offield, T.W., Abdullah, S.K. and Ali, S.T., 1975. Geology of the southern Himalaya in Hazara, Pakistan, and adjacent areas. 716C.
Chen, T., Zhu, L., Niu, R.-q., Trinder, C.J., Peng, L. and Lei, T., 2020. Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. Journal of Mountain Science, 17(3): 670-685.
Chen, W., Chen, X., Peng, J., Panahi, M. and Lee, S., 2021. Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer. Geoscience Frontiers, 12(1): 93-107.
Chen, W., Li, W., Hou, E., Zhao, Z., Deng, N., Bai, H. and Wang, D., 2014. Landslide susceptibility mapping based on GIS and information value model for the Chencang District of Baoji, China. Arabian Journal of Geosciences, 7(11): 4499-4511.
Chen, W., Sun, Z. and Han, J., 2019. Landslide Susceptibility Modeling Using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. Applied Sciences, 9(1).
Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z. and Hong, H., 2017. GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. Geomatics, Natural Hazards and Risk, 8(2): 950-973.
Ciurleo, M., Cascini, L. and Calvello, M., 2017. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. Engineering Geology, 223: 71-81.
Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1): 37-46.
Das, I., Sahoo, S., van Westen, C., Stein, A. and Hack, R., 2010. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass

classification system, along a road section in the northern Himalayas (India). Geomorphology, 114(4): 627-637.
Di Napoli, M., Carotenuto, F., Cevasco, A., Confuorto, P., Di Martire, D., Firpo, M., Pepe, G., Raso, E. and Calcaterra, D., 2020. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. Landslides, 17(8): 1897-1914.
Ding, Q., Chen, W. and Hong, H., 2017. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. Geocarto International, 32(6): 619-639.
Dou, J., Tien Bui, D., P. Yunus, A., Jia, K., Song, X., Revhaug, I., Xia, H. and Zhu, Z., 2015. Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. PLOS ONE, 10(7): e0133262.
Dou, J., Yunus, A.P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., Avtar, R., Chen, Y., Pham, B.T. and Yamagishi, H., 2020. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Science of The Total Environment, 720: 137320.
Dou, J., Yunus, A.P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Khosravi, K., Yang, Y. and Pham, B.T., 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Science of The Total Environment, 662: 332-346.
Du, G.-l., Zhang, Y.-s., Iqbal, J., Yang, Z.-h. and Yao, X., 2017. Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. Journal of Mountain Science, 14(2): 249-268.
Erener, A. and Düzgün, H.S.B., 2012. Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? Environmental Earth Sciences, 66(3): 859-877.
Fallah-Zazuli, M., Vafaeinejad, A., Alesheykh, A.A., Modiri, M. and Aghamohammadi, H., 2019. Mapping landslide susceptibility in the Zagros Mountains, Iran: a comparative study of different data mining models. Earth Science Informatics, 12(4): 615-628.
Franek, J. and Kresta, A., 2014. Judgment Scales and Consistency Measure in AHP. Procedia Economics and Finance, 12: 164-173.
Frattini, P., Crosta, G. and Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Engineering Geology, 111(1): 62-72.
Gardezi, S.A.H., Hussain, G., Neupane, B., Imran, M., Hamid, Q.Y., Ikram, N., Usmani, N.A. and Asghar, H., 2021. Geological Investigation of 5.6 MW Mirpur Earthquake, Northwestern Himalayas, Pakistan. International Research Journal of Earth Sciences, 09(01): 20-31.
Gerrard, J., 1994. The landslide hazard in the Himalayas: geological control and human action. Geomorphology, 10(1): 221-230.
Goetz, J.N., Guthrie, R.H. and Brenning, A., 2011. Integrating physical and empirical landslide susceptibility models using generalized additive models. Geomorphology, 129(3): 376-386.
Guzzetti, F., 2005. Landslide Hazard and Risk Assessment. PhD Thesis, 389 pp pp.
Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.-T., 2012. Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, 112(1): 42-66.
Hong, H., Chen, W., Xu, C., Youssef, A.M., Pradhan, B. and Tien Bui, D., 2017. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using

frequency ratio, certainty factor, and index of entropy. Geocarto International, 32(2): 139-154.
Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.X., Chen, W. and Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). CATENA, 163: 399-413.
Hong, H., Tsangaratos, P., Ilia, I., Loupasakis, C. and Wang, Y., 2020. Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. Science of The Total Environment, 742: 140549.
Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S. and Guo, Z., 2020. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. CATENA, 191: 104580.
Hussain, A. and Khan, R.N., 1996. Geological map of Azad Jammu and Kashmir. Geological Survey of Pakistan, Geological Map Series.
Hussain, A., Yeats, R.S., Kausar, A.B., Karim, T. and Khan, T., 2006. The Balakot–Bagh fault that triggered the October 8, earthquake and other active faults in the Himalayan foreland region, Pakistan, International Conference on Earthquake in Pakistan: Extended Abstracts. Geological Survey of Pakistan, Quetta, pp. 115-116.
Hussain, A., Yeats, R.S. and MonaLisa, 2009. Geological setting of the 8 October 2005 Kashmir earthquake. Journal of Seismology, 13(3): 315-325.
Hussain, G., Fang, X., Usmani, N.A., Gardezi, S.A.H., Hussain, M., Asghar, H., Paryal, M. and Khalid, S., 2020. Structural and stratigraphic studies of Hazara-Kashmir Syntaxis, northwestern Himalaya, Pakistan. North American Academic Research, 03(09): 1-14.
Jadoon, I.A.K., Hinderer, M., Kausar, A.B., Qureshi, A.A., Baig, M.S., Basharat, M. and Frisch, W., 2015. Structural interpretation and geo-hazard assessment of a locking line: 2005 Kashmir earthquake, western Himalayas. Environmental Earth Sciences, 73(11): 7587-7602.
James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. Statistical Learning. In: G. James, D. Witten, T. Hastie and R. Tibshirani (Editors), An Introduction to Statistical Learning: with Applications in R. Springer New York, New York, NY, pp. 15-57.
Kamp, U., Growley, B.J., Khattak, G.A. and Owen, L.A., 2008. GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. Geomorphology, 101(4): 631-642.
Kaneda, H., Nakata, T., Tsutsumi, H., Kondo, H., Sugito, N., Awata, Y., Akhtar, S.S., Majid, A., Khattak, W., Awan, A.A., Yeats, R.S., Hussain, A., Ashraf, M., Wesnousky, S.G. and Kausar, A.B., 2008. Surface Rupture of the 2005 Kashmir, Pakistan, Earthquake and Its Active Tectonic Implications. Bulletin of the Seismological Society of America, 98(2): 521-557.
Khan, S.F., Kamp, U. and Owen, L.A., 2013. Documenting five years of landsliding after the 2005 Kashmir earthquake, using repeat photography. Geomorphology, 197: 45-55.
Khattak, G.A., Owen, L.A., Kamp, U. and Harp, E.L., 2010. Evolution of earthquake-triggered landslides in the Kashmir Himalaya, northern Pakistan. Geomorphology, 115(1): 102-108.
Kira, K. and Rendell, L.A., 1992. A Practical Approach to Feature Selection. In: D. Sleeman and P. Edwards (Editors), Machine Learning Proceedings 1992. Morgan Kaufmann, San Francisco (CA), pp. 249-256.
Kononenko, I., 1994. Estimating attributes: Analysis and extensions of RELIEF. In: F. Bergadano and L. De Raedt (Editors), Machine Learning: ECML-94. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 171-182.

Lever, J., Krzywinski, M. and Altman, N., 2016. Logistic regression. Nature Methods, 13(7): 541-542.
Li, L., Lan, H., Guo, C., Zhang, Y., Li, Q. and Wu, Y., 2017. A modified frequency ratio method for landslide susceptibility assessment. Landslides, 14(2): 727-741.
Liu, J. and Duan, Z., 2018. Quantitative Assessment of Landslide Susceptibility Comparing Statistical Index, Index of Entropy, and Weights of Evidence in the Shangnan Area, China. Entropy, 20(11).
Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R. and Abderrahmane, B., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207: 103225.
Murtagh, F., 1991. Multilayer perceptrons for classification and regression. Neurocomputing, 2(5): 183-197.
Nefeslioglu, H.A., Duman, T.Y. and Durmaz, S., 2008. Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey). Geomorphology, 94(3): 401-418.
Nettleton, D., 2014. Chapter 6 - Selection of Variables and Factor Derivation. In: D. Nettleton (Editor), Commercial Data Mining. Morgan Kaufmann, Boston, pp. 79-104.
Owen, L.A., Kamp, U., Khattak, G.A., Harp, E.L., Keefer, D.K. and Bauer, M.A., 2008. Landslides triggered by the 8 October 2005 Kashmir earthquake. Geomorphology, 94(1): 1-9.
Park, S., Choi, C., Kim, B. and Kim, J., 2013. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environmental Earth Sciences, 68(5): 1443-1464.
Pourghasemi, H.R. and Rahmati, O., 2018. Prediction of the landslide susceptibility: Which algorithm, which precision? CATENA, 162: 177-192.
Pourghasemi, H.R., Teimoori Yansari, Z., Panagos, P. and Pradhan, B., 2018. Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). Arabian Journal of Geosciences, 11(9): 193.
Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences, 51: 350-365.
Pradhan, B. and Jebur, M.N., 2017. Spatial Prediction of Landslide-Prone Areas Through k-Nearest Neighbor Algorithm and Logistic Regression Model Using High Resolution Airborne Laser Scanning Data. In: B. Pradhan (Editor), Laser Scanning Applications in Landslide Assessment. Springer International Publishing, Cham, pp. 151-165.
Quinlan, J.R., 1993. C 4.5: Programs for machine learning. The Morgan Kaufmann Series in Machine Learning.
Rahman, A.-u., Khan, A.N. and Collins, A.E., 2014. Analysis of landslide causes and associated damages in the Kashmir Himalayas of Pakistan. Natural Hazards, 71(1): 803-821.
Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T. and Akgun, A., 2014. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arabian Journal of Geosciences, 7(2): 725-742.
Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Reviews, 180: 60-91.

Rusk, J., Maharjan, A., Tiwari, P., Chen, T.-H.K., Shneiderman, S., Turin, M. and Seto, K.C., 2022. Multi-hazard susceptibility and exposure assessment of the Hindu Kush Himalaya. Science of The Total Environment, 804: 150039.
Saaty, T.L., 1990. How to make a decision: The analytic hierarchy process. European Journal of Operational Research, 48(1): 9-26.
Saba, S.B., van der Meijde, M. and van der Werff, H., 2010. Spatiotemporal landslide detection for the 2005 Kashmir earthquake region. Geomorphology, 124(1): 17-25.
Saha, A. and Saha, S., 2020. Comparing the efficiency of weight of evidence, support vector machine and their ensemble approaches in landslide susceptibility modelling: A study on Kurseong region of Darjeeling Himalaya, India. Remote Sensing Applications: Society and Environment, 19: 100323.
Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7): 1308.
Sana, H. and Nath, S.K., 2017. Seismic source zoning and maximum credible earthquake prognosis of the Greater Kashmir Territory, NW Himalaya. Journal of Seismology, 21(2): 411-424.
Searle, M.P. and Treloar, P.J., 2019. Introduction to Himalayan tectonics: a modern synthesis. Geological Society, London, Special Publications, 483(1): 1.
Shafique, M., 2020. Spatial and temporal evolution of co-seismic landslides after the 2005 Kashmir earthquake. Geomorphology, 362: 107228.
Shafique, M., van der Meijde, M. and Khan, M.A., 2016. A review of the 2005 Kashmir earthquake-induced landslides; from a remote sensing prospective. Journal of Asian Earth Sciences, 118: 68-80.
Shah, S.M.I., 2009. Stratigraphy of Pakistan. Geologyical Survey of Pakistan, 22: 2539.
Shahabi, H. and Hashim, M., 2015. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Scientific Reports, 5(1): 9899.
Shahabi, H., Khezri, S., Ahmad, B.B. and Hashim, M., 2014. Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. CATENA, 115: 55-70.
Sun, X., Chen, J., Han, X., Bao, Y., Zhan, J. and Peng, W., 2020. Application of a GIS-based slope unit method for landslide susceptibility mapping along the rapidly uplifting section of the upper Jinsha River, South-Western China. Bulletin of Engineering Geology and the Environment, 79(1): 533-549.
Taud, H. and Mas, J.F., 2018. Multilayer Perceptron (MLP). In: M.T. Camacho Olmedo, M. Paegelow, J.-F. Mas and F. Escobar (Editors), Geomatic Approaches for Modeling Land Change Scenarios. Springer International Publishing, Cham, pp. 451-455.
Tien Bui, D., Pradhan, B., Lofman, O. and Revhaug, I., 2012. Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. Mathematical Problems in Engineering, 2012: 974638.
Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B. and Revhaug, I., 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(2): 361-378.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2): 127-150.
Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S. and Moore, J.H., 2018. Relief-based feature selection: Introduction and review. Journal of Biomedical Informatics, 85: 189-203.

Wang, H., Zhang, L., Yin, K., Luo, H. and Li, J., 2021. Landslide identification using machine learning. Geoscience Frontiers, 12(1): 351-364.
Wang, K., Xu, H., Zhang, S., Wei, F. and Xie, W., 2020a. Identification and Extraction of Geomorphological Features of Landslides Using Slope Units for Landslide Analysis. ISPRS International Journal of Geo-Information, 9(4).
Wang, Q., Li, W., Wu, Y., Pei, Y. and Xie, P., 2016. Application of statistical index and index of entropy methods to landslide susceptibility assessment in Gongliu (Xinjiang, China). Environmental Earth Sciences, 75(7): 599.
Wang, Y., Fang, Z., Wang, M., Peng, L. and Hong, H., 2020b. Comparative study of landslide susceptibility mapping with different recurrent neural networks. Computers & Geosciences, 138: 104445.
Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J. and Steinberg, D., 2008. Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1): 1-37.
Yalcin, A., Reis, S., Aydinoglu, A. and Yomralioglu, T.J.C., 2011. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. CATENA, 85(3): 274-287.
Zare, M., Pourghasemi, H.R., Vafakhah, M. and Pradhan, B., 2013. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arabian Journal of Geosciences, 6(8): 2873-2888.
Zhang, T., Han, L., Chen, W. and Shahabi, H., 2018. Hybrid Integration Approach of Entropy with Logistic Regression and Support Vector Machine for Landslide Susceptibility Modeling. Entropy, 20(11).

Statements and Declarations

Funding

It is declared that any author has NOT received any funding from any person or organization.

Conflicts of interest

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

Authors' contributions

Conceptualization: Syed Ahsan Hussain Gardezi and Nadeem Ahmad Usmani;

Methodology: Syed Ahsan Hussain Gardezi, Nadeem Ahmad Usmani and Nawaz Ikram;

Formal analysis and investigation: Syed Ahsan Hussain Gardezi, Nadeem Ahmad Usmani,

Saad Wani Writing - original draft preparation: Syed Ahsan Hussain Gardezi and Nadeem

Ahmad Usmani; Writing - review and editing: Nawaz Ikram and Sajjad Ahmad; Supervision: Xiao-qing Chen and Sajjad Ahmad

Ethics approval

It is stated that the work described has not been published previously, is not considered elsewhere and approved by all authors. It is also stated that the manuscript is self-written to avoid plagiarism, and no exaggeration in results and other data has been made. Also, none of the citation used in the manuscript is unnecessary.

Consent for publication

Will be provided when demanded.