Application of frequency ratio, information value, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Murree region, Sub-Himalayas

Fakhrul Islam  
COMSATS University Islamabad (CUI)

Muhammad Farooq Iqbal  
(✉️ farooq buzdar@gmail.com )  
COMSATS University Islamabad (CUI)

Irfan Mahmood  
COMSATS University Islamabad (CUI)

Muhammad Imran Shahzad  
COMSATS University Islamabad (CUI)

Safeer Ullah Shah  
Ministry of Climate Change

Research Article

Keywords: Landslide Susceptibility, Geospatial techniques, Bivariate Models, RUSLE Model, Sub Himalayas

Posted Date: November 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2218881/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Landslide is a chronic geohazard in hilly regions which affects the socioeconomic trends globally. Current study is conducted to apply three models including Information Value (IV), Frequency Ratio (FR) and Weights of Evidence (WoE) for Landslide Susceptibility Mapping (LSM) of Murree. Sentinel-2, Google Earth, and field surveys data were used to generate an inventory map of 102 landslides and these events were divided into two subsets i.e., 70% and 30% for LSM and model validation respectively. Eleven causative factors including soil erosion, elevation, slope, aspect, curvature, drainage, fault, road, precipitation, Land Use Land Cover (LULC), and lithology maps were prepared using Google Earth Engine (GEE). The final susceptibility maps were produced with the training datasets of landslide events and causative factors using IV, FR and WoE, whereas these maps were validated using the Receiver Operating Characteristic (ROC) technique. The Area Under Curve (AUC) illustrated the Success Rate Curve (SRC) of 69%, 78% and 79% for the IV, FR and WoE models, respectively, while Predicted Rate Curve (PRC) were 80%, 95% and 87% for the IV, FR and WoE models, respectively. The results of this study can be used by policymakers to plan some mitigation regarding soil erosion and landslides-prone region.

Introduction

Landslide is a frequent destructive geohazard phenomenon in the mountainous regions (Chen et al. 2016). Landslide is downslope movement of a geological body mass due to instability of slope by some internal and external stresses in the area. Landslide is caused by anthropogenic and natural processes. The anthropogenic landslides-induced factors are slope re-profiling, water networking, land-use changes, artificial structures, and mining, whereas natural factors are geologic, climatic, hydrologic, and topographic. Landslides produce environmental issues in the hilly region including but not limited to soil degradation and deforestation (Van et al. 2017). This hazard has particularly affected the demographic trend of Sub-Himalayan region of Pakistan (Saleem et al. 2020).

Qualitative, semi-quantitative and quantitative techniques are used to design the precise Landslide Susceptibility Mapping (LSM) (Zezere et al. 2017). The qualitative techniques were widely used by geo-experts in the late 1970 era. The qualitative technique consists of inventory, analysis of geomorphology, empirical problem-solving experience, and evaluation expertise in the relevant field. The qualitative techniques are cost-effective; however, the output results are not most authentic in landslide investigations (Shano et al. 2020). In the recent epoch, with the advancement of technology, qualitative techniques updated to semi quantitative techniques including multi-criteria decision support system for landslide investigation, but the results were still not satisfactory for investigation due to less accuracy (Feizizadeh and Blaschke 2014). In the modern era quantitative techniques like bivariate and multivariate geospatial modelling emerged as an authentic and reliable techniques for LSM. The quantitative technique is an objective based technique, where association of landslide events and predisposing factors is processed objectively through statistical, deterministic, and probabilistic ways to generate LSM. Furthermore, quantitative techniques are updated with geospatial approaches including but not limited to Geographical Information System (GIS), Global Positioning System (GPS) and satellite based – Remote Sensing (RS) datasets. Currently, the most widely used method by the researchers and technical experts’ teams for the LSM and hazard assessment are the GIS and RS based quantitative approaches which were utilized in the current study.

The LSM demarcates the potential areas inclined to landslides and classifies the region into different landslides predicted zones. The evaluation of risk associated with landslides is a persistent goal of many eras for geoscientists and geological engineers, but it became feasible and expressible due to the accessibility of RS data (Sharma et al. 2014). The RS data can be helpful to produce inventory of events and preparation of causative parameters in LSM using GIS techniques (Shahabi and Hashim 2015). The hilly region of China is mostly affected by landslide activities therefore, GIS based techniques were utilized to produce LSM in northeastern China to mitigate and prevent the hazard (Yu and
LSM by using geospatial techniques helped the policy makers of India to mitigate the hazard in eastern Himalayas (Chawla et al. 2018). GIS based statistical models were performed in China to produce authentic LSM of the area (Chen et al. 2016).

LSM has been predicted many susceptible zones of landslides in Pakistan (Khan et al. 2013). Landslide is the most common and destructive environmental hazard in the northern regions of Pakistan (Khan et al. 2019). LSM along China Pakistan Economic Corridor (CPEC) was performed to identify sensitive zones of landslides and to mitigate its impacts regarding socioeconomic development (Ali et al. 2018). LSM of eastern zone of Pakistan was generated in 2019 using geospatial techniques to identify landslide prone regions (Gilany 2019). All causative factors have significant association with landslide hazards but the most important factor in causing landslides is soil erosion in the area which has been mostly ignored by researchers in Pakistan. Currently, soil erodibility is the major geohazard of modern time that affects the environment (Pradhan et al. 2012). The most widely used model for soil erosion is the Revised Universal Soil Loss Equation (RUSLE). The priority by the researcher to RUSLE is due to some adequate reasons like it is easier to implement, understandable functions, and its compatibility with the GIS environment (Huang et al. 2016).

It is imperative to propose appropriate techniques for soil erosion mapping and LSM to overcome the damages of geohazard in the study area to minimize the consequences of landslides and soil erosion and, to protect the environment from the mentioned hazards in mountainous regions (Chen et al. 2016). The goal of current study is to develop landslide inventory map and to assess the performance of Information Value (IV), Frequency Ratio (FR) and Weight of Evidence (WoE) bivariate models for LSM computation and to analyze landslide events with soil erosion and all other causative factors in the Murree area of Rawalpindi district. Furthermore, Area Under Curve Receiver Operating Characteristic (AUROC) technique is used to validate the performance of IV, FR and WoE bivariate models.

Materials And Methods

Study Area

Current research work is conducted in Murree District, as shown in Fig. 1. Geographically, research area is situated at latitude and longitude range of 33° 45' 0" N and 73° 33' 0" E respectively with elevation of 2300 meters (Satti et al. 2017). Geologically, study area lies at the foothills of Western Himalayas. The direction of the strata in the study area dips towards the core of the ridge which develops resistance to the instability of strata in the area. Mass wasting in the region is a common hazard due to unconsolidated material in different geological formations (Abbasi et al. 2002). Most of the area is dominated by lithological beds of sandstone, siltstone, clay of Murree and Kuldana formation which were deposited in Oligocene to Miocene age. The study area also has calcareous and shale beds. The area has weak lithological beds on sheer steep slopes due to which landslides initiated in the area (Ahmed et al. 2020).

Datasets

The datasets utilized for LSM development include both ground and satellite RS data. The ground data consists of precipitation, geological, soil and road data. Daily Surface Precipitation Gauge (SPG) data (mm/day) was acquired from Pakistan Meteorological Department (PMD), whereas geological map was acquired from Geological Survey of Pakistan (GSP). Soil data was acquired from Soil Survey of Pakistan and road data was acquired from the Punjab Highway Authority (PHA). The RS data utilized consists of Digital Elevation Model (DEM), Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), Landsat 8 Operational Land Imager (OLI), Google Earth, and Sentinel-2 datasets. DEM data was used to compute elevation, slope, aspect, curvature and drainage network. Spatial resolution of CHIRPS is 0.05° (~ 5.54 Km) and daily gridded data were used for the precipitation assessment. Landsat 8 OLI data with
a spatial resolution of 30 meters was used for the Land Use Land Cover (LULC) mapping of the study area. Sentinel-2 data for 2015, 2016, 2017, 2018, 2019, and 2020 were integrated with the field data of landslides and Google Earth Imageries in GIS to produce landslide inventory maps of the study area.

**Methodology**

**Landslide Inventory Map**

Landslide inventory is the first step for LSM to predict future landslide hazards in the region (Ding et al. 2017). Sentinel-2 satellite data and Google Earth imageries were utilized to generate landslides inventory maps of the study area. Sentinel-2 and Google Earth based updated inventory maps were developed and validated using field surveys data. By utilizing this technique, most updated inventory maps were developed.

**Predisposing Factors**

The causative factors were prepared and evaluated using Google Earth Engine (GEE) and GIS environment to understand their association with landslides. GEE and GIS are advanced and improved platforms for the researchers to understand the relationship of landslides with soil erosion in the area of interest (Pradhan et al. 2012). GIS environment was utilized to compute the soil erosion factors including precipitation factor (R-factor), soil erodibility factor (K-factor), slope length and slope steepness (LS-factor), and cover management factor (C-factor). The numerical estimation of the parameters in the region of concern is shown in Eq. 1.

\[
A = R \times K \times LS \times C \times P \tag{1}
\]

Where, \(A\) is the soil loss (tons/ha/yr), \(R\) is Precipitation erosivity factor (MJmm/ ha.hr.yr), \(K\) is soil erodibility factor (tons/ha/yr), \(LS\) is slope length and steepness factor (Dimensionless), \(C\) is cover management factor (dimensionless) and \(P\) is conservation practice factor (dimensionless). The R-factor was computed using Eq. 2 (Singh et al. 2010).

\[
R = 79 + 0.363 \times P \tag{2}
\]

The soil erodibility factor (\(K\)) significantly influenced the soil erosion (Ganasri and Ramesh 2016). The parameter \(L\) and \(S\) combinedly show the impacts of slope length and slope steepness on soil erosion (Saleem et al. 2018). The topographic factor was computed using Eq. 3.

\[
LS = \sqrt{L \times \frac{R}{22.1}} (0.065 + 0.045 \times S + 0.0065 \times S^2) \tag{3}
\]

Where, \(L\) is the flow length or flow accumulation; \(R\) is the Resolution and \(S\) is slope. The cover management factor (\(C\)) of the model is used to examine the impacts of different cover to soil erosion in the area (Chadli 2016). Elevation is a crucial internal factor in landslides vulnerability assessment in the area (Cellek 2020). In the present study, elevation was computed from Shuttle Radar Topography Mission (SRTM) DEM using GEE. Slope is an important parameter in downslope movement of strata (Cellek 2020). The angle of slope is very important in landslide studies (Pradhan et al. 2017). Aspect is the direction of topography which alters the moisture contents and vegetation surface in the area initiates the weathering process (Silalahi et al. 2019). Curvature describes the inclination of slope. The concave structure is more susceptible to landslides than the convex surface and flat (Xu 2014). The drainage network was computed from DEM to understand their impacts on slope instability in the region of interest.

The precipitation creeps into pore spaces of soil and initiates the instability in the region due to saturation between grains (Hong et al. 2017). The annual mean precipitation map was prepared from data of CHIRPS for 2015 to 2019 in
GEE and exported to the GIS environment. Faults are an influential factor in landslides study. The geological fault map was derived from the geological map of Northern Pakistan (Searle and Khan, 2001). Lithological factors are influential parameters in landslide investigation and every bed has different susceptibility to landslides due to their composition, moisture and mechanical properties (Chen et al. 2019). The road network is another important factor in the instability of strata in hilly regions which leads to landslides in the area (Chen et al. 2019).

**Information Value (IV)**

The IV is a bivariate model used to develop the association of landslides with its predisposing factors in the region to produce LSM of study area (Wubalim et al. 2020). This model as shown in Eq. 4 guides about the role of each parameter with respect to landslides events.

\[
W = \log \frac{\frac{N_{p|x}(S_i)}{N_{p|x}(N_i)}}{\sum \frac{N_{p|x}(S_i)}{\sum N_{p|x}(N_i)}}
\]

Where, \(W\) shows the weight of class of each predisposing factor for landslides. \(N_{p|x}(S_i)\) denotes the number of landslides pixels within class “i”, \(N_{p|x}(N_i)\) number of all pixels within class “i”, \(N_{p|x}(S_i)\) total number of landslides pixels, \(\sum N_{p|x}(N_i)\) is used for total number of pixels in study area. After utilizing the above equation for weight of each factor class, Eq. 5 was used for computation of LSM.

\[
LS = W_S + W_A + W_{LULC} + W_L + W_P + W_F + W_R + W_D + W_C + W_E + W_S
\]

Where, \(W_S\) is the weight of slope, \(W_A\) is the weight of aspect, \(W_{LULC}\) is the weight of LULC, \(W_L\) is the weight of lithology, \(W_S\) is the weight of soil, \(W_P\) is the weight of fault, \(W_R\) is the weight of road, \(W_P\) is the weight of precipitation and \(W_D\) is the weight of drainage.

**Frequency Ratio Model**

The second method for landslide susceptibility assessment is the Frequency Ratio Model. This model is a bivariate model utilized for LSM development of the study area using geospatial techniques (Carter et al. 1994). Eq. 6 was used for computation of frequency ratio for landslides susceptibility.

\[
FR = \frac{N_i Q_x / N}{N_i L Q / N L}
\]

Where, \(FR\) is the frequency ratio, \(N_i Q_x\) is the number of pixels in each landslide's conditioning factor class, \(N\) is the number of all pixels in study area, \(N_i L Q\) is the number of landslides pixels in each landslide conditioning factors \(N L\) is the number of all landslide pixels in study area. Eq. 7 was used for computation of LSM map from frequency ratio method.

\[
LS = \sum_{i=1}^{n} FR_{ij}
\]

The above expression \(FR_{ij}\) is frequency ratio value for ‘j’ class of factor ‘i’, \(n\) is the total number of factors.

**Weight of Evidence**
The third model for LSM is Weight of Evidence which is also a bivariate statistical method. This model is working on the principle of Bayesian theorem approach developed to compute the probability based on the concept of prior and posterior probability (Getachew and Meten, 2021). The causative factors of landslides were used as an input parameter in this model to deliver the association of landslide's events with these factors and spatial location extent of landslides. Weight of evidence modelling was performed using the GIS environment. Eqs. (8) and (9) were used to assess predisposing factors influence on LS in the GIS environment.

\[ W^+ = \ln \frac{NP_i x_1}{NP_i x_2} + \frac{NP_i x_2}{NP_i x_3} + \frac{NP_i x_3}{NP_i x_4} \]  
\[ W^- = \ln \frac{NP_i x_3}{NP_i x_1} + \frac{NP_i x_2}{NP_i x_4} + \frac{NP_i x_3}{NP_i x_4} \]

Where, \( NP_i x_1 \) is the number of pixels representing the presence of both landslide causative factor and landslides, \( NP_i x_2 \) is the absence of landslides causative factor and presence of landslides, \( NP_i x_3 \) is the present of landslides causative factor and absence of landslides, whereas \( NP_i x_4 \) is the absence of both landslides and landslides causative factor.

Results

Around 102 landslide events were detected using satellite-based RS techniques and filed surveys. Figure 1 displays the spatial locations of the detected landslide events including debris flow (20), flow slide (24), rockfall (16) and slide (42) for the studied period. The RUSLE model parameters were utilized to generate soil erosion maps. The relationship of inventory and predisposing parameters computed the occurrence of events in any class of all parameters in percentage. Figure 2a displays the soil erodibility map generated using soil map. Texture of soil is a very influential factor in soil erodibility and value of soil erodibility ranges from 0.22 to 0.33 Mg h MJ-1mm-1 for loamy and non-calcareous soil, and loamy and clayey non calcareous soil respectively. Figure 2b displays the mean precipitation erosivity factor in mm/day. Figure 2c displays the slope length and steepness and Fig. 2d displays the cover management factor computed from LULC. Figure 2e displays the final generated soil erosion map and it was prepared using soil erodibility factor (K), precipitation erosivity factor (R), slope length and steepness (LS) and cover management factor (C) parameters. Soil erosion map is classified into four zones including low, medium, high, and very high. Table 1 shows the results of major factors for controlling the landslides in the area. It was observed that landslide is directly linked with the soil erosion in the study area and the increasing soil erosion is more susceptible to landslide occurrences.
Table 1
Detailed analysis of different causative parameters with landslide events using Bivariate Models (IV, FR and WoE) to understand the impacts of each class in LSM.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Class</th>
<th>No of pixels in class</th>
<th>No of landslide pixels in a class</th>
<th>W+</th>
<th>W−</th>
<th>WC</th>
<th>% Pixels in Class</th>
<th>% LS pixels in Class</th>
<th>(FR)</th>
<th>IV = log (A/B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil Erosion</td>
<td>Low</td>
<td>64767</td>
<td>215</td>
<td>-0.69</td>
<td>0.07</td>
<td>-1.07</td>
<td>13.93</td>
<td>6.18</td>
<td>0.44</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>155318</td>
<td>1071</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.06</td>
<td>33.41</td>
<td>30.81</td>
<td>0.92</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>191446</td>
<td>1380</td>
<td>0.10</td>
<td>-0.01</td>
<td>0.11</td>
<td>41.18</td>
<td>39.70</td>
<td>0.96</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Very High</td>
<td>53350</td>
<td>390</td>
<td>0.08</td>
<td>-0.06</td>
<td>0.15</td>
<td>11.47</td>
<td>11.21</td>
<td>0.97</td>
<td>0.09</td>
</tr>
<tr>
<td>Elevation</td>
<td>&lt; 900</td>
<td>107632</td>
<td>1203</td>
<td>0.45</td>
<td>-0.10</td>
<td>0.35</td>
<td>22.46</td>
<td>34.57</td>
<td>1.53</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>900–1100</td>
<td>71920</td>
<td>817</td>
<td>0.77</td>
<td>-0.17</td>
<td>0.60</td>
<td>15.01</td>
<td>23.48</td>
<td>1.56</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>1100–1300</td>
<td>85410</td>
<td>1158</td>
<td>0.103</td>
<td>-0.20</td>
<td>0.83</td>
<td>17.82</td>
<td>33.28</td>
<td>1.86</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>1300–1500</td>
<td>72886</td>
<td>118</td>
<td>-1.50</td>
<td>0.13</td>
<td>-1.63</td>
<td>15.21</td>
<td>3.39</td>
<td>0.22</td>
<td>-1.50</td>
</tr>
<tr>
<td></td>
<td>&gt; 1500</td>
<td>141251</td>
<td>183</td>
<td>-1.72</td>
<td>0.29</td>
<td>-2.02</td>
<td>29.48</td>
<td>5.26</td>
<td>0.17</td>
<td>-1.72</td>
</tr>
<tr>
<td>Slope</td>
<td>&lt; 10</td>
<td>60344</td>
<td>134</td>
<td>-1.19</td>
<td>0.09</td>
<td>-1.29</td>
<td>12.70</td>
<td>3.86</td>
<td>0.30</td>
<td>-1.18</td>
</tr>
<tr>
<td></td>
<td>10–20</td>
<td>217254</td>
<td>1419</td>
<td>-0.11</td>
<td>0.08</td>
<td>-0.19</td>
<td>45.73</td>
<td>40.97</td>
<td>0.89</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>20–30</td>
<td>165432</td>
<td>1590</td>
<td>0.33</td>
<td>-0.02</td>
<td>0.36</td>
<td>34.82</td>
<td>45.91</td>
<td>1.31</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>30–40</td>
<td>30547</td>
<td>310</td>
<td>0.27</td>
<td>-0.18</td>
<td>0.46</td>
<td>6.43</td>
<td>8.95</td>
<td>1.39</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>&gt; 40</td>
<td>1461</td>
<td>10</td>
<td>-0.06</td>
<td>0.00</td>
<td>-0.06</td>
<td>0.30</td>
<td>0.28</td>
<td>0.93</td>
<td>-0.06</td>
</tr>
<tr>
<td>Aspect</td>
<td>F</td>
<td>36821</td>
<td>206</td>
<td>-0.26</td>
<td>0.01</td>
<td>-0.28</td>
<td>7.75</td>
<td>5.94</td>
<td>0.76</td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>NE</td>
<td>34546</td>
<td>218</td>
<td>-0.14</td>
<td>0.01</td>
<td>-0.15</td>
<td>7.27</td>
<td>6.29</td>
<td>0.86</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>52413</td>
<td>581</td>
<td>0.42</td>
<td>-0.06</td>
<td>0.49</td>
<td>11.03</td>
<td>16.77</td>
<td>1.52</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>74371</td>
<td>842</td>
<td>0.44</td>
<td>-0.10</td>
<td>0.55</td>
<td>15.65</td>
<td>24.31</td>
<td>1.55</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>63720</td>
<td>60</td>
<td>-2.05</td>
<td>0.12</td>
<td>-2.18</td>
<td>13.41</td>
<td>1.73</td>
<td>0.12</td>
<td>-2.04</td>
</tr>
<tr>
<td></td>
<td>SW</td>
<td>42743</td>
<td>94</td>
<td>-1.20</td>
<td>0.06</td>
<td>-1.27</td>
<td>8.99</td>
<td>2.71</td>
<td>0.30</td>
<td>-1.198</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>47930</td>
<td>425</td>
<td>0.19</td>
<td>-0.02</td>
<td>0.22</td>
<td>10.08</td>
<td>12.27</td>
<td>1.21</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>NW</td>
<td>61501</td>
<td>689</td>
<td>0.43</td>
<td>-0.08</td>
<td>0.51</td>
<td>12.94</td>
<td>19.89</td>
<td>1.53</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>60993</td>
<td>348</td>
<td>-0.24</td>
<td>0.03</td>
<td>-0.27</td>
<td>12.83</td>
<td>10.04</td>
<td>0.78</td>
<td>-0.24</td>
</tr>
<tr>
<td>Curvature</td>
<td>Concave</td>
<td>118738</td>
<td>1031</td>
<td>0.18</td>
<td>-0.06</td>
<td>0.24</td>
<td>24.78</td>
<td>29.63</td>
<td>1.19</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Flat</td>
<td>262259</td>
<td>1909</td>
<td>0.002</td>
<td>0.002</td>
<td>0.005</td>
<td>54.74</td>
<td>54.87</td>
<td>1.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Convex</td>
<td>98102</td>
<td>539</td>
<td>-0.28</td>
<td>0.06</td>
<td>-0.34</td>
<td>20.47</td>
<td>15.49</td>
<td>0.75</td>
<td>-0.27</td>
</tr>
<tr>
<td>Precipitation</td>
<td>1410–1571</td>
<td>47185</td>
<td>1096</td>
<td>1.17</td>
<td>-0.27</td>
<td>1.45</td>
<td>9.93</td>
<td>31.64</td>
<td>3.18</td>
<td>1.15</td>
</tr>
<tr>
<td>Parameters</td>
<td>Class</td>
<td>No of pixels in class</td>
<td>No of Landslide pixels in a class</td>
<td>( W^+ )</td>
<td>( W^- )</td>
<td>WC</td>
<td>% Pixels in Class</td>
<td>% LS pixels in Class</td>
<td>(FR)</td>
<td>IV = ( \log \frac{A}{B} )</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------</td>
<td>-----------------------</td>
<td>-----------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----</td>
<td>------------------</td>
<td>---------------------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>1571 – 1681</td>
<td>98970</td>
<td>1146</td>
<td>0.46</td>
<td>-0.16</td>
<td>0.63</td>
<td>20.83</td>
<td>33.09</td>
<td>1.58</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>1681 – 1773</td>
<td>94900</td>
<td>764</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.12</td>
<td>19.97</td>
<td>22.06</td>
<td>1.10</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>1773 – 1881</td>
<td>72489</td>
<td>167</td>
<td>-1.15</td>
<td>0.11</td>
<td>-1.27</td>
<td>15.25</td>
<td>4.82</td>
<td>0.31</td>
<td>-1.15</td>
</tr>
<tr>
<td></td>
<td>1881 – 2035</td>
<td>161494</td>
<td>290</td>
<td>-1.40</td>
<td>0.33</td>
<td>-1.73</td>
<td>33.99</td>
<td>8.37</td>
<td>0.24</td>
<td>-1.40</td>
</tr>
<tr>
<td>LULC</td>
<td>Forest</td>
<td>210650</td>
<td>860</td>
<td>-0.57</td>
<td>0.29</td>
<td>-0.87</td>
<td>43.96</td>
<td>24.79</td>
<td>0.56</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>Vegetation</td>
<td>141362</td>
<td>919</td>
<td>-0.10</td>
<td>0.042</td>
<td>-0.15</td>
<td>29.50</td>
<td>26.49</td>
<td>0.89</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>Barren Land</td>
<td>63576</td>
<td>1188</td>
<td>0.96</td>
<td>-0.27</td>
<td>1.23</td>
<td>13.26</td>
<td>34.25</td>
<td>2.58</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>61743</td>
<td>496</td>
<td>0.10</td>
<td>-0.01</td>
<td>0.12</td>
<td>12.88</td>
<td>14.30</td>
<td>1.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>1796</td>
<td>5</td>
<td>-0.95</td>
<td>0.00</td>
<td>-0.96</td>
<td>0.37</td>
<td>0.14</td>
<td>0.38</td>
<td>-0.95</td>
</tr>
<tr>
<td>Distance to Fault</td>
<td>&lt; 25</td>
<td>2468</td>
<td>11</td>
<td>-0.49</td>
<td>0.00</td>
<td>-0.49</td>
<td>0.51</td>
<td>0.31</td>
<td>0.61</td>
<td>-0.48</td>
</tr>
<tr>
<td></td>
<td>25 – 75</td>
<td>5000</td>
<td>27</td>
<td>-0.29</td>
<td>0.002</td>
<td>-0.30</td>
<td>1.04</td>
<td>0.77</td>
<td>0.74</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>75 – 150</td>
<td>7488</td>
<td>43</td>
<td>-0.23</td>
<td>0.003</td>
<td>-0.23</td>
<td>1.56</td>
<td>1.23</td>
<td>0.79</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>150 – 350</td>
<td>19894</td>
<td>215</td>
<td>0.40</td>
<td>-0.02</td>
<td>0.42</td>
<td>4.15</td>
<td>6.17</td>
<td>1.48</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>&gt; 350</td>
<td>444248</td>
<td>3183</td>
<td>-0.01</td>
<td>0.15</td>
<td>-0.17</td>
<td>92.72</td>
<td>91.49</td>
<td>0.98</td>
<td>-0.01</td>
</tr>
<tr>
<td>Lithology</td>
<td>Alluvium</td>
<td>1269</td>
<td>9</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.26</td>
<td>0.25</td>
<td>0.97</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>Quaternary</td>
<td>18192</td>
<td>116</td>
<td>0.002</td>
<td>-0.001</td>
<td>-0.47</td>
<td>3.79</td>
<td>3.33</td>
<td>0.87</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>Muree Formation</td>
<td>398228</td>
<td>2898</td>
<td>-0.13</td>
<td>0.04</td>
<td>0.61</td>
<td>83.12</td>
<td>83.29</td>
<td>1.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Kuldana</td>
<td>20835</td>
<td>340</td>
<td>0.81</td>
<td>-0.05</td>
<td>0.87</td>
<td>4.34</td>
<td>11.52</td>
<td>2.24</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Lora</td>
<td>10620</td>
<td>70</td>
<td>-2.15</td>
<td>0.09</td>
<td>-2.06</td>
<td>2.21</td>
<td>0.25</td>
<td>0.90</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>Margalla Hill</td>
<td>6774</td>
<td>30</td>
<td>-1.87</td>
<td>0.04</td>
<td>-1.65</td>
<td>5.58</td>
<td>0.86</td>
<td>0.15</td>
<td>-1.86</td>
</tr>
<tr>
<td></td>
<td>Limestone</td>
<td>683</td>
<td>15</td>
<td>1.12</td>
<td>-0.002</td>
<td>1.60</td>
<td>0.14</td>
<td>0.43</td>
<td>3.02</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Shale</td>
<td>697</td>
<td>1</td>
<td>-1.62</td>
<td>0.001</td>
<td>-1.62</td>
<td>0.14</td>
<td>0.02</td>
<td>0.19</td>
<td>-1.62</td>
</tr>
<tr>
<td>Distance to Road</td>
<td>&lt; 20</td>
<td>6437</td>
<td>92</td>
<td>0.68</td>
<td>-0.01</td>
<td>0.69</td>
<td>1.34</td>
<td>0.63</td>
<td>1.96</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>20 – 40</td>
<td>6401</td>
<td>64</td>
<td>0.32</td>
<td>-0.005</td>
<td>0.32</td>
<td>1.33</td>
<td>0.51</td>
<td>1.37</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>40 – 100</td>
<td>18535</td>
<td>157</td>
<td>0.15</td>
<td>-0.006</td>
<td>0.16</td>
<td>3.86</td>
<td>1.55</td>
<td>1.16</td>
<td>0.15</td>
</tr>
<tr>
<td>Parameters</td>
<td>Class</td>
<td>No of pixels in class</td>
<td>No of Landslide pixels in a class</td>
<td>W^+</td>
<td>W^-</td>
<td>WC</td>
<td>% Pixels in Class</td>
<td>% LS pixels in Class</td>
<td>(FR)</td>
<td>IV = log (A/B)</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>----------------------------------</td>
<td>-------</td>
<td>------</td>
<td>-----</td>
<td>------------------</td>
<td>----------------------</td>
<td>------</td>
<td>--------------</td>
</tr>
<tr>
<td>100–350</td>
<td>67753</td>
<td>516</td>
<td>0.04</td>
<td>-0.008</td>
<td>0.05</td>
<td>14.14</td>
<td>3.93</td>
<td>1.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>&gt;350</td>
<td>379972</td>
<td>2650</td>
<td>-0.04</td>
<td>0.14</td>
<td>-0.18</td>
<td>79.30</td>
<td>93.36</td>
<td>0.96</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>Distance to Stream</td>
<td>&lt; 25</td>
<td>6968</td>
<td>145</td>
<td>1.06</td>
<td>-0.02</td>
<td>1.09</td>
<td>1.45</td>
<td>4.16</td>
<td>2.86</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>25–50</td>
<td>9840</td>
<td>201</td>
<td>1.04</td>
<td>-0.03</td>
<td>1.08</td>
<td>2.053</td>
<td>5.77</td>
<td>2.81</td>
<td>1.006</td>
</tr>
<tr>
<td></td>
<td>50–100</td>
<td>18071</td>
<td>359</td>
<td>1.01</td>
<td>-0.07</td>
<td>0.94</td>
<td>3.77</td>
<td>10.31</td>
<td>2.73</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>100–250</td>
<td>47475</td>
<td>800</td>
<td>0.85</td>
<td>-0.15</td>
<td>1.009</td>
<td>9.90</td>
<td>22.99</td>
<td>2.32</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td>&gt;250</td>
<td>396744</td>
<td>1974</td>
<td>-0.38</td>
<td>0.93</td>
<td>-1.31</td>
<td>82.81</td>
<td>56.74</td>
<td>0.68</td>
<td>-1.05</td>
</tr>
</tbody>
</table>

Figure 3 displays the topographic parameters extracted using DEM data for developing LSM map, where Fig. 3a displays the elevation map. Table 1 shows a strong association of landslide events with elevation up to 1300 meter, whereas decrease of landslide events occurrence was observed with the increasing elevation. Figure 3b displays the slope in degree map and it was noted that slope is an active parameter in landslides occurrence. Table 1 shows that up to 45-degree slope, increased landslide events were observed with the increase of slope in the study area. Figure 3c displays the aspect map. Table 1 shows that the class SE was more susceptible to landslides in the study area followed by NW and E direction. These three classes were major contributors in landslides occurrence in the study area. Figure 3d displays the curvature map classified into concave, at and convex. Every class of curvature has different impacts to landslides susceptibility. Table 1 shows that concave structures are more susceptible to landslides because in concave structure, the materials are accumulated which leads to sliding in the area whereas in convex the loose materials in the study area are removed and only compacted lithology was left on extreme convex structures.

Figure 4 displays the parameters other than topography for LSM generation including distance to stream, precipitation, LULC, distance to fault, lithology and distance to roads were also mapped. Figure 4a displays that stream buffers were applied for drainage to understand their association with landslide events. To investigate LSM, different sizes of buffers including 250m were applied in bivariate models with landslide events. Table 1 shows that the relationship of every class was different with landslide events. The results of all models revealed that the drainage network was an influential factor in landslides occurrences. A well-developed strong association between both variables was observed.

Figure 4b displays an annual mean precipitation (mm/year) map prepared using CHIRPS satellite-based data for the period from 2015 to 2019. Precipitation was the most influential external triggering factor of landslide occurrences. CHIRPS precipitation data was validated with SPG prior to utilizing the data for analysis. Point-to-point comparison shows CC of 0.79 and 0.87 whereas point-to-grid shows CC of 0.80 and 0.87 on monthly and annual timescales respectively. The MAE and RMSE were higher on point-to-point analysis as compared to point-to-grid analysis on a monthly timescale, however, MAE and RMSE were almost the same on annual timescale.

Table 1 shows that precipitation has the highest value range as compared to the rest of parameters for bivariate models as 1.15, 3.18 and 1.45 respectively. The results revealed that 1410–5171 mm/year precipitation class was more triggering for landslides for the current study followed by 5171–1681 and 1681–1773 mm/year respectively. The results explained that precipitation was a crucial factor for LSM. Figure 4c displays satellite based LULC classified maps including forest, vegetation cover, barren land, urban and water bodies. LULC was also an influential parameter in LSM, because every LULC element has different impacts in hazards generation processes. Table 1 shows that some elements
mitigate landslide occurrences whereas other elements facilitate the slope instability in the region. It was observed that the barren land class of LULC was a crucial factor in landslide occurrence in the study area which was followed by anthropogenic activities in the form of construction in the area. Both classes were active in LSM.

Figure 4d displays the tectonic map of faults prepared in the GIS environment and reclassified into five buffered zones along with the lineament. Table 1 shows the association of different buffer zones of tectonic faults with landslide events. Five fault buffers were developed ranging from < 25m to > 350m and associated these classes with events by IV, FR and WoE. The results revealed that 150 to 350 classes are more prone and susceptible to landslides. Figure 4e displays the lithology developed using a geological map, which was scanned, digitized, and extracted according to the study area. Table 1 shows that lithology was an essential internal controlling factor for LSM. The association of different lithological units like Alluvium, Quaternary, Murree Formation, Kuldana Formation, Lora Formation, Margalla Hill limestone, Kuzagali Shales and Mari Limestone events performed by IV, FR and WoE. The results revealed that Kuzagali shales are more susceptible to landslides and followed by Kuldana and Murree Formation. Figure 4f displays buffers along the road network which were digitized from Google Earth and validated with the ground data. Road distances were reclassified into 5 classes ranging from < 20m to > 350m to develop association of this parameter with landslide activities using IV, FR and WoE.

Table 1 shows that the class of Fig. 5 displays the final LSM map developed using three different bivariate models including IV, WoE and FR using Eq. 4 to 9. Figure 5a displays the final LSM developed using the IV model, Fig. 5b displays the LSM map developed using FR, whereas Fig. 5c displays the final LSM map developed using the WoE model. All the three bivariate models were validated, so researchers may choose the best model for their research work (Abraham et al. 2020).

The AUROC technique was utilized to validate the performance of three bivariate models opted in the study including IV, FR and WoE. Figure 6 displays the validation results of IV, FR and WoE models based on 30% of inventory data of landslide events. The AUC graph for IV model revealed the graph values 0.69 and 0.80 for SRC and PRC respectively. These values can be mentioned in percentage values for model accuracy assessment which are 69% and 80% for training data and validation data respectively for LSM. The SR and PR for IV model is shown in Fig. 6a and 6b.

The AUC graph for FR model explained that the graph values 0.78 and 0.95 for SR and PR respectively. These values can be mentioned in percentage values for model accuracy assessment which are 78% and 95% for training data and validation data respectively for LSM. The SRC and PRC for FR model is shown in Fig. 6c and 6d. The AUC graph for WoE model exposed the graph values 0.79 and 0.87 for SR and PR respectively. These values can be mentioned in percentage values for model accuracy assessment which are 79% and 87% for training data and validation data respectively for LSM. The SRC and PRC for WoE model is shown in Fig. 6e and 6f. The validation results of all these models clearly revealed that FR was the best model for the LSM of the current research area.

Discussion

For LSM development, eleven predisposing factors (soil erosion, elevation, slope, aspect, curvature, precipitation, LULC, distance to fault, lithology, distance to road and distance to streams) were considered and prepared using GEE and GIS platforms. These factors were adequate to understand geohazards in the study area. Three Bivariate models including IV, FR and WoE were also utilized to investigate the parameters for LSM. All the parameters were reclassified according to the literature.

Soil erosion is strongly correlated with landslide occurrences (Pradhan et al. 2012). It was observed that a very high class of soil erosion was more susceptible to landslides and vice versa. The most soil erosion class has the highest numbers, however, the low eroded area has lowest value for IV, FR, and WoE respectively. The elevation data also played
a pivotal role for landslides detection and investigation. One major factor was the association of increasing landslide events with increasing altitude. Similar results were also observed at an elevation range of 1500 to 600m by Nakileza et al. (2020). Schurz et al. (2019) also observed maximum landslide events in the elevation range of 1500 to 1800m. In the current study, elevation parameters ranging from 1100 to 1300 m had the highest bivariate analysis results such as 0.62, 1.86 and 0.83 for IV, FR and WoE respectively.

The results of this study clearly explained that more landslide events occurred in the slope angle up to 45 degrees. The landslides occurrence also increases which shows the strong association of slope classes with landslides events. Yu et al. (2020) also observed similar results that slope angles between 30° to 45° are more susceptible to landslide events. The positive association was not developed above 45-degree slope because the current research area is dominantly occupied by unconsolidated materials which eroded completely up to 45°. So, there was no loose material left in the region to be eroded, and only hard rocks were left in the area resistant to erosion. The results revealed that the slope classes between 20 to 30° and 30 to 40° are very susceptible to landslides. It was also observed that the aspect class of SE is more vulnerable to landslides in the study area. The more susceptible class of aspect is SE with IV, FR and WoE ranges for this class are 0.44, 1.55 and 0.55 respectively.

The concave structures are more susceptible to landslides because, in concave structures, the materials can be accumulated, leading to sliding in the area. In convex, the loose material in the currents study area is removed with only compacted lithology left on outer convex structures (Xu, 2014). In the current study, it was observed that concave structure is the most susceptible zone to landslides than other structures. The literature illustrated that precipitation is the most influential external triggering and catastrophic causative factor of landslides (Jeong et al. 2017; Abraham et al. 2020). In this study, precipitation has the highest value range as compared to the rest of parameters for bivariate models such as 1.15, 3.18 and 1.45 for IV, FR and WoE models respectively.

The barren land is the more susceptible class of LULC to landslide with FR value of 2.82 (Khan et al. 2019). Similar results were also observed in the current study, where the barren land class of LULC has a significant contribution to landslide event occurrences. The FR value for barren land in this study was observed as 2.58. The results illustrated that the 150 to 350 m class of fault is more susceptible to landslide. The IV, FR and WoE models range for 150 to 350 m were 0.39, 1.48 and 0.42, respectively. The landslide susceptibility is incomplete without incorporating the geological information like fold, fault, stress, hydrogeology, and lithological composition (Igwe 2015). The more susceptible lithological formation to landslides was Kuzagali Shale followed by the lithology of Kuldana and Murree Formations. These units are unconsolidated in nature.

The proper management of road design is very important to reduce the risk of slope failure and slope instability. The construction of road networks in hilly regions is always considered the most human induced landslide factor because geological strata lead to triggering of hazard. Landslides have crucial impacts on road channels and effects the demographic trend of the area (Donnini et al. 2017). The detection and recognition of susceptibility zones along roadside is extremely essential to reduce the causality and financial loss due to potential hazard in the area (Bordoni et al. 2018). The most susceptible class of road parameters were < 20 mm. In the current study, the road network was strongly influenced by the occurrence of landslides.

The stream network is another influential hydrologic factor of landslide with 1.55 and 0.62 for < 25m of FR and WoE respectively (Pradhan et al. 2012). In current research stream networks also showed a good association with landslide events. The more susceptible class of stream network parameter was also < 25m zone. The result illustrated that the value of FR and WoE for < 25m zone are 2.81 and 1.06 respectively. The above discussion concluded that the highest value for LSM in the current research area by bivariate models was the value of precipitation, lithology, and stream
network. The validation results revealed that FR model is best model for the LSM because the model showed 78% and 95% accuracy for SRC and PRC, respectively.

Conclusions

The findings of this research work delivered key information regarding landslide occurrence in the study area. The output results revealed that the landslides were caused by different predisposing factors in the region. Current study was performed to identify the zones susceptible to landslides in the study region to mitigate their consequences using GIS and RS. The base of this work analysis was on the association of landslide events with predisposing factors. These parameters were topographic, climatic, and geomorphic. The bivariate statistical analysis was performed for the research work to evaluate the association of dependent and independent variables for LSM.

This study generated landslide susceptibility assessment in Sub Himalayas by integrating geospatial technology and field data. Data of Sentinel-2, Google Earth, and field surveys were used to generate an inventory map of 102 detected landslide events in the study area (debris flow, flow slide, rockfall and slide).

Eleven parameters were considered as a crucial factor for landslide susceptibility mapping. Three models were used in this study to compute the effects of each parameter in landslide susceptibility. The results were validated to compare the performance of models. The validated results revealed that the FR model is the best model for the LSM of the current research area with 78% and 95% for SRC and PRC respectively. The results illustrated that precipitation and lithology have a significant role in landslide occurrence followed by drainage network, barren land, road network, elevation, aspect, and slope. It can be concluded that geospatial modelling is a more reliable technique to produce LSM. The final LSM of this study can be used by organizations to mitigate the soil erosion and landslide hazard in the region.

Declarations

Acknowledgments

The authors would like to thank Pakistan Meteorological Department (PMD), Punjab Highway Authority, (PHA) Soil Survey of Pakistan (SSP) and Geological Survey of Pakistan (GSP) for providing ground data. We would like to thank Climate Hazards Center University of California, Santa Barbara for CHIRPS data. We would like to thank the European Space Agency (ESA) for Sentinel-2 data. and United State Geological Survey (USGS) for Landsat 8 data.

Funding

No Funding was involved.

Competing Interests

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 paper.

Availability of data and materials

Open-source datasets were utilized in this study, however, field date is available from the corresponding author on reasonable request.

Author Contributions
Fakhrul Islam and Muhammad Farooq Iqbal has designed the research methodology, collected data, and analyzed, discussed, and wrote the manuscript. Irfan Mahmood has made significant contribution in data analysis, interpretation and writing this manuscript. Muhammad Imran Shahzad and Safeer Ullah Shah contributed for performing analysis and shaping this manuscript.

References


**Figures**
Figure 1

Map of the study area along with elevation in meters and spatial location of landslide events.
Figure 2

RUSLE model parameters including (a) Soil erodibility factor, (b) Precipitation erosivity factor, (c) Slope length and steepness, (d) Cover management factor and (e) Soil erosion map.
Figure 3

Topographic parameters derived from DEM for LSM including (a) Elevation, (b) Slope, (c) Aspect and (d) Curvature map.
Figure 4

Different parameters for Landslide Susceptibility Mapping include (a) Stream buffer, (b) Precipitation, (c) LULC, (d) Fault buffer, (e) Lithology and (f) Road buffer.
Figure 5

Landslide Susceptibility Map derived from Bivariate models using (a) LSM by IV (b) LSM by FR and (c) LSM by WoE.
Figure 6

Success Rate Curve (SRC) and Predicted Rate Curve (PRC) for IV, FR, and WoE models, where (a) shows SRC for IV, (b) shows PRC for IV, (c) shows SRC for FR Model, (d) shows PRC for FR Model, (e) shows SRC for WoE and (f) shows PRC for WoE.