Research Article

Evaluation of Shannon Entropy and Weights of Evidence Models in Landslide Susceptibility Mapping for the Pithoragarh District of Uttarakhand State, India

Nguyen Duc Dam,1 Mahdis Amiri,2 Nadhir Al-Ansari3, Indra Prakash4, Hiep Van Le1, Hanh Bich Thi Nguyen1, and Binh Thai Pham1

1University of Transport Technology, 54 Trieu Khuc, Than Xuan, Hanoi 100000, Vietnam
2Department of Watershed & Arid Zone Management, Gorgan University of Agricultural Sciences & Natural Resources, Gorgan 4918943464, Iran
3Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden
4DDG (R) Geological Survey of India, Gandhinagar 382010, India

Correspondence should be addressed to Nadhir Al-Ansari; nadhir.alansari@ltu.se and Binh Thai Pham; binhopt@utt.edu.vn

Received 16 December 2020; Revised 27 April 2021; Accepted 26 February 2022; Published 12 April 2022

Academic Editor: Domenico Magisano

Copyright © 2022 Nguyen Duc Dam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Landslide susceptibility mapping is considered a useful tool for planning, disaster management, and natural hazard mitigation of a region. Although there are different methods for predicting landslide susceptibility, the bivariate statistical analysis method is considered to be simple and popular. In this study, the main aim is to evaluate the performance of Shannon entropy (SE) and weights of evidence (WOE) statistical models in landslide susceptibility mapping of Pithoragarh district of Uttarakhand state, India. For this purpose, ten landslide affecting factors, namely, slope degree, aspect, curvature, elevation, land cover, slope forming materials, geomorphology (landforms), distance to rivers, distance to roads, and overburden depth were used for the development of landslide susceptibility maps using the SE and WOE methods. Data extracted from the Google Earth images, Aster Digital Elevation Model, and Geological Survey of India report were used for the construction and evaluation of landslide susceptibility models and maps. The landslide data of 91 locations were randomly divided into two parts in the ratio of 70:30 using GIS software that is 70% data was used for training the models and 30% data was used for testing and validating the models. Performance of the applied models was evaluated using area under the AUC (area under the curve) ROC (receiver operating characteristics) curve. Results indicated that the WOE model is having better accuracy (AUCWOE = 68.75%) than the SE model (AUCSE = 52.17%) in the development of landslide susceptibility maps. Hence, WOE model can be used for the development of accurate landslide susceptibility maps which can provide useful information to decision maker and policy planner in better development of landslide prone areas.

1. Introduction

Landslides are one of the most dangerous and threatening natural disasters for various communities around the world causing huge loss of life, economy, and infrastructures [1]. Landslides are downward gravitational movement of rocks and soil [1]. Landslides destabilizes the ground stability and can enhance soil formation in the area [2, 3]. It can also affect sustainable development goals of the United Nations and the land degradation neutrality challenge [4]. Factors which affect landslides include geology, topography, land use pattern, rainfall and other geo-environmental factors. Out of these, geology is one of the most important factors in the occurrence of landslides [5]. Recently, climate change effects have increased landslides events all over the world [6, 7].

Landslide susceptibility map is an important tool in the landslide management of the area [8, 9]. These maps help in the reduction of landslide risk on human life, infrastructures,
communication and in proper planning of land use [10]. Landslide susceptibility refers to the probability of a landslide which may occur in an area in future, based on the past events under similar conditions [11]. There are various methods of generation of landslide susceptibility maps such as statistical methods, in conjunction with Geographical Information System (GIS) and remote sensing techniques [12–16]. Some of these methods include multivariate regression, likelihood ratio, information value, logistic regression, and binary logistic regression [17–22]. Other algorithms which have acceptable accuracy in landslide susceptibility mapping are discriminant analysis, analytic hierarchy processes, weights of evidence, weighted linear combinations, evidential belief functions, and generalized additive models [8, 23–25]. Shannon’s entropy is another important bivariate algorithm used to analyze landslide susceptibility [26]. Other statistical methods for landslide susceptibility mapping (zoning) are probabilistic models, certainty factors, information values, modified bayesian estimation, weighting variables, weighted linear combinations of instability variables, landslide nominal risk variables [27–30], spatial multicriteria evaluation, index of entropy, and Dempster–Shafer models [31–36].

In recent years, researchers have widely used machine learning (ML) algorithms in the natural hazard studies including landslide susceptibility modeling such as support vector machine, multivariate adaptive regression spline, boosted regression, classification and regression trees, Naïve Bayes, quadratic discriminant analysis, artificial neural networks, maximum entropy, random forest, and generalized linear model [37–44]. However, simple statistical methods/models are also being applied in the landslide studies to understand relationship between affecting variables/factors and occurrence of landslides. In this study, we have used Shannon entropy (SE) and weights of evidence (WOE) popular statistical models for the development of landslide susceptibility maps of Pithoragarh district of Uttarakhand State, India, which is one of the prominent landslide prone areas in the Himalayan region. For the evaluation of statistical models, we have used ten landslide affecting factors: slope degree, aspect, curvature, and elevation, land cover, slope forming materials (SFM), geomorphology (landforms), distance to rivers, distance to roads, and overburden depth. Rainfall is generally acted as a triggering factor in the occurrence of landslides in the Himalayan region, therefore this factor has not been considered separately in the model studies. Data for the use in statistical models were extracted from the Geological Survey of India (GSI) report (https://www.gsi.gov.in/webcenter/portal), Google Earth images, and Aster Digital Elevation Model (DEM). Performance of the models was evaluated using area under the receiver operating characteristic (ROC) curve method. GIS software was used for data integration and visualization.

2. Study Area

Topography of the Pithoragarh district is rugged mountainous mark by steep hills and deep valleys (Figure 1). General elevation in the study area ranges from 1500 m (in south) to 2500 m (in north) above mean sea level. Sarju and Ramganga Rivers are prominent rivers traversing the area. Google Earth images and GSI map shows that landslides are very prominent along the excavated slopes of road sections and on steep sides of river valleys such as Bansura-Rameshwar Ghat Road and along Sarju river (Figure 2). Geologically, rocks of almora crystalline (granitoids) and garhwal group (shale, slate, phylite, quartzite, dolomite, limestone, magnesite, occasional calc slate, and metavolcanics) separated by thrust fault are present in this area (https://www.gsi.gov.in/webcenter/portal). The area is affected by tectonic activities indicated by folded and faulted rocks. Quaternary sediments are present in river valleys, on hill slopes and as glacial deposits.

Climate of the area varies from the moist lower elevation to cold temperate and rain shadow (higher elevation). There are four main seasons in the district: winter (December to mid-March), summer (mid-March to mid June), and rainy (mid-June to mid-September), transitional that is season of retreating monsoon (mid-September to November). Temperature varies from subzero in winter (at higher elevation) to 40 to 45 degree in summer (at lower elevation).

Most part of the Pithoragarh district is a highly vulnerable to landslides, hence selected as study area. The area is tectonically disturbed and geologically dissected by faults, shears, and joints and thus affected by more landslides in the event of heavy rains or due to excavation of roads and other infrastructure projects. Numerous landslides occurrences have been recorded in the past and present necessitating systematic studies for proper monitoring and prevention of landslides for the proper development of the area. The most catastrophic landslide caused by unprecedented rains, killing 221 people occurred in the early morning of August 18, 1998, in Malpa in Pithoragarh district [45].

3. Materials and Methods

Methodological flowchart of this study is presented in Figure 3 which is self-explanatory:

3.1. Shannon Entropy. Entropy is one of the management approaches that has been used to deal with the disorder, instability, turbulence, and uncertainty in a system and has shown the amount of uncertainty in a continuous probability distribution [46]. In fact, the entropy is a concept that estimates dispersion and disorder in natural phenomena. This concept, which is widely used in the science of thermodynamics, has spread to other sciences today [47]. The theory of entropy was first quantified by Shannon [46]. In Shannon entropy (SE) [46], the variables with the maximum impact on the occurrence of an event are determined [48]. To use this model, a decision matrix must first be created. The decision matrix contains information that entropy can be used to evaluate, and by calculating the entropy matrix and the total weight of the ten landslide factors used in this study. The following equations were used to plot the map using the SE model [27, 48]:
**Figure 1:** Landslide's location map of study area.

**Figure 2:** GSI Photographs of landslides in the study area (source: Geological Survey of India report: https://www.gsi.gov.in/webcenter/portal).
Where \( P_{ij} \) is the landslide density in each class, \( (P_{ij}) \) is the probability of landslides in each factor and its related classes, \( H_j \) and \( (H_j) \) are entropy values and maximum entropy, \( I_j \) is the information coefficient, \( W_{ij} \) is the final weight of each factor. After determining the final weight of each factor and multiplying it in the mentioned factor classes, i.e., the \( P_{ij} \) values of each factor and its classes based on dividing the number of landslides by the number of pixels of the factor classes, finally the weight maps were added together. Final landslide maps were prepared and the more sensitive floors took on more weight [49].

3.2. Weights of Evidence. Weights of evidence (WOE) or conditional probability method was first developed to identify and explore mineral deposits [50]. Conditional probability analysis is a valuable tool in determining risk zoning especially, when appropriate factors and good knowledge of the variables affecting the landslide are available. This theory is a data-driven method that is known as one of the models of Bayesian theory in the form of linear algorithm and uses the previous (unconditional) and posterior (conditional) probabilities [51]. This method (WOE) is used when sufficient data is considered to estimate the relative importance of the issues substantiated through statistics. For this purpose, first, the probabilistic weights of each factor and its related classifications are calculated using the following relationships [52].

\[
W^+_i = \log_e \left( \frac{P(B_i|S)}{P(B|S)} \right), \quad W^-_i = \log_e \left( \frac{P(B|S)}{P(B_i|S)} \right).
\]
dicting landslide susceptibility has a value of AUC 0.5 and 1 [57]. The model with excellent accuracy in precision of the models [57]. The AUC value varies between prediction maps [56]. Area under ROC curve, detects the X-axis (1-specificity) with different cut-off, which was used positive rate on dimension graph that has landslide susceptibility as real (true) spatial modeling process [55]. The ROC curve is a two dimensional characteristic (ROC) curve is a well-known method for
characteristic (ROC) Curve.

3.3. Evaluation Method Using Receiver Operating Characteristic (ROC) Curve. Area under the receiver operating characteristic (ROC) curve is a well-known method for evaluating and comparing the accuracy of algorithms used to prepare landslide susceptibility maps [54]. Validation and accuracy evaluation of bivariate statistical models were done on 30% testing or validation data which was randomly selected from the landslide occurrence points during the spatial modeling process [55]. The ROC curve is a two

dimension graph that has landslide susceptibility as real (true) positive rate on Y-axis (sensitivity) and false positive rate on X-axis (1-specificity) with different cut-off, which was used for numerical appraisal of landslide susceptibility of the prediction maps [56]. Area under ROC curve, detects the precision of the models [57]. The AUC value varies between 0.5 and 1 [57]. The model with excellent accuracy in predicting landslide susceptibility has a value of AUC = 1 and the weak (non-instructive) model has an AUC of 0.5. As well as, the area below the ROC curve represents the predicted value of the system by describing its ability to accurately estimate the occurrence or occurrence of landslide events and the absence of landslides [57].

3.4. Data Used

3.4.1. Landslide Inventory. One of the most important part of landslide susceptibility mapping and zoning is to prepare landslide inventory showing geospatial distribution of landslide events in a map [58]. For the development of models, landslide polygon data represented by points on the map was randomly divided into two parts in the ratio of 70:30 as training dataset and validation data set, respectively [59–61]. The training data set (70%) was used for landslide susceptibility mapping/zoning and remaining 30% testing dataset (validation) for the validation and accuracy evaluation of the models used.

In this study, the inventory of past 91 landslide events was prepared based on the available past record of Geological Survey of India (https://www.gsi.gov.in/webcenter/portal) which showed a predominant percentage of debris slip (61 cases) and less number of rock slips (17 cases) and slight falls/flips (3 cases). Furthermore, there were only 6 cases of deep damage (5 debris slip, 1 rock slip) while the rest were shallow damage. In addition, several recent landslides were included in the inventory by interpreting of Google Earth images (Table 1, Figures 1 and 2). Most of the landslides in the area occurred during heavy rainfall especially along roads and on steep hill/valley slopes covered by loose soil and rock debris and in the areas where ground-mass/rock mass is weathered and dissected by structural discontinuities.

3.4.2. Factors Affecting Landslides. There are still significant differences of opinion on the selection of important variables having impact on landslides [62]. Important features in the selection of influencing factors should have easy access and trustworthy accuracy [63]. In this study, ten affecting factors namely slope degree, aspect, curvature, elevation, land cover, Slope Forming Materials (SFM), geomorphology (landforms), distance to rivers, distance to roads, and overburden depth were considered based on the local geo-environmental conditions. Thematic layers of the factors from Aster DEM were generated with 30-pixel cell size using GIS software. Slope degree, aspect, curvature, elevation maps were prepared from the Aster DEM map of the region downloaded from USGS (https://earthexplorer.usgs.gov) and other layers were extracted from the available geological, geomorphological, land cover maps obtained from Geological Survey of India reports (https://www.gsi.gov.in/webcenter/portal) and Google Earth images (Table 1). Data used in this study are also presented in Tran, Dam, Jalal, Al-Ansari, Ho, Phong, Iqbal, Le, Nguyen and Prakash [64].

(1) Slope Degree. Slope is one of the important and effective factors on the occurrence of landslides. The reason why the slope variable is important in landslide susceptibility assessment is that it controls surface and subsurface flow and directly affects runoff and infiltration [65]. In the present study, a slope angle map in five classes was prepared from DEM using natural break method of GIS software (Figure 4).

(2) Aspect. The slope aspect is considered as an important parameter in assessing landslide susceptibility [66] as the sun, air/wind and rain/precipitation of the region affect in different directions [67]. On the other hand, aspect indirectly affects the vegetation and soil moisture. In the present study, the slope aspect map is divided into nine classes: flat, north, northeast, east, southeast, south, southwest, west and northwest using DEM (Figure 4).

(3) Curvature. Curvature of the slope plays an important role in the surface run off and ground infiltration thus affects the
Table 1: Data used in this study.

<table>
<thead>
<tr>
<th>No</th>
<th>Data</th>
<th>Source</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Landslide inventory</td>
<td>Geological Survey of India report</td>
<td>1:50,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>google earth images</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Digital elevation model (DEM)</td>
<td>USGS (<a href="https://earthexplorer.usgs.gov">https://earthexplorer.usgs.gov</a>)</td>
<td>30m</td>
</tr>
<tr>
<td>3</td>
<td>Land cover</td>
<td>Geological Survey of India report</td>
<td>1:50,000</td>
</tr>
<tr>
<td>4</td>
<td>Landforms (geomorphology)</td>
<td>Geological Survey of India report</td>
<td>1:50,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Google Earth images</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Roads</td>
<td>DEM (<a href="https://earthexplorer.usgs.gov">https://earthexplorer.usgs.gov</a>)</td>
<td>30m</td>
</tr>
<tr>
<td>6</td>
<td>Slope forming materials (SFM)</td>
<td>Geological Survey of India report</td>
<td>1:50,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Google Earth images</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Continued.
erosion of the surface and ground water condition of the region [66]. Thus, this factor affects the occurrence of landslides. The curvature map of this area was prepared from DEM and classified in the concave, convex and flat surfaces using GIS software (Figure 4).

(4) Elevation. Elevation is an important factor in the occurrence of landslides [66]. In general, landslides occur in the hilly areas. At higher elevations generally rains are less but glaciers are prominent. Most of the rains and vegetation is confined at lower and middle elevations in Himalayas. Landslides events depend on the elevation where slopes are moderate to high with heavy rainfall and less vegetation. The study area was divided into nine classes of slopes from DEM using natural break method of Arc GIS (Figure 4).

(5) Land Cover. In general, bare lands and unvegetated areas are more prone to landslides than lands with dense vegetation cover and forested areas [68]. In the highly vegetated area roots of plants act as a reinforcement of the ground and prevent erosion. In the dense forest area impact of the rains directly on the ground is very less due to foliage and thus less erosion. Landslides also occur in the cultivated area due to percolation of water during irrigation and also due to erosion of top soil. The land use map of the study area was extracted from the data available from Geological Survey of India report (https://www.gsi.gov.in/webcenter/portal). The land use map of the study area consists of eleven main groups (Figure 4).

(6) Slope Forming Materials (SFM). The SFM map was extracted from the data available on the Geological Survey of India website (https://www.gsi.gov.in/webcenter/portal). Type of the SFM is very important in the shallow landslide study. Characteristics of the groundmass depend on the SFM; its permeability, porosity and geotechnical properties. Landslides depend on the above characteristics of the material and also on the size and binding/looseness of soil and joints in the rock mass. The SFM map of the present study area was classified into eighteen main groups (Figure 4).

(7) Geomorphology (Landforms). Geomorphology which is a study of landforms is an important factor in the study of landslide susceptibility [18]. Geomorphological features such as mountains, valleys, river terraces, undulating grounds, ridges and escarpments etc., and affects the occurrence of landslides in conjunction with other topographic and geo-environmental factors. The relevant landform...
Factors were extracted from the data available on the Geological Survey of India report (https://www.gsi.gov.in/webcenter/portal) (Figure 4).

(8) Distance to Rivers and Roads. Distance from roads is one of the anthropogenic factors used in landslide susceptibility assessment. Presence of road network and absence of roads in the area affect landslide occurrences [69]. Roadside excavation and vegetation removal are activities that cause landslides during road construction [43]. Anthropogenic activities such as excavation of roads create instability of slopes near and adjacent to roads up to certain distances depending on the nature of ground mass and geology. Hence, the distance from the road is a very impressive factor in the landslide study [26]. Similarly, distance to rivers also plays an important role in the assessment of landslides for the development of landslide susceptibility maps. The hydrological network regime, soil saturation of water sources, and groundwater recharge, as well as increasing water pressure to empty water pores, lead to landslides in areas adjacent to water sources, rivers and streams [70]. Road distance and river distance buffer maps were prepared for the landslide susceptibility mapping (Figure 4).

(9) Overburden Depth. Chances of slope failure are more likely in thick overburden areas depending on the characteristics of the overburden material. Major part of the study area is covered thin over burden (1–3 m) with occasional pockets of greater than 5 m. Thus, the possibility of landslides due to failure of overburden material is very less. However, nature and thickness of material affects the infiltration and thus ground water conditions in the area which may affect moisture conditions of the underlying rock mass creating instability and thus landslides. Overburden material map was extracted from the data available on the Geological Survey of India report (https://www.gsi.gov.in/webcenter/portal) (Figure 4).

4. Results and Discussion

4.1. Analysis of WOE and SE Models Results. Table 2 shows the results of landslide susceptibility analysis of landslide affecting factors using the WOE and SE models. The slope analysis results show that the highest number of landslides (60) occurrence are in slope range 41.57°–75.19° in case of both the models. The difference between positive and negative weights (C) in the WOE method and landslide density (Pij) in the SE method in the moderate to high slope is the highest. As shown by decreasing the degree of weight gradient the presence of landslides decreased. Generally, on the lower slopes of the resistant forces are more than the driving forces and the condition of the landslide occurrence is not favourable.

The maximum number of landslides is in the direction of the west aspect with 50 and the highest numerical value of C = 0.94 and Pij = 0.27 based on the WOE and SE ways is related to the west aspect, respectively. After that, in the southwest, southeast, and south directions of the study area, the highest number of landslides (44, 34, and 32) has been seen, which can be attributed to climatic conditions such as high humidity, which is consistent with the results of other research [27].

The effect of landforms variable on landslide susceptibility in the case area shows that the highest number of landslides on the moderately dissected hills slope formation occurred due to large area, but according to the WOE and SE methods, the highest weight was related to "Piedmont slope" formation with C = 1.32 and the density (Pij) is 0.25. Also, in the Colluvial foot slope and "Denudational hillslope" formations, there was the highest landslide weight C and density (Pij), and conversely, there was no landslide susceptibility on the "River" class. In general, it can be stated that rock areas are more resistant to weathering and dispersal so landslides are less likely to occur [71]. But, the geological formations of each region are unique, so the results of the relationship between landslides occur and in different geological formations are various and specific to each region.

The results of curvature (topographic morphology) show that the highest landslide susceptibility is related to concave slopes and convex slopes has less susceptibility and finally flat slopes have the lowest susceptibility. Typically, the positive curvature of a convex topography is upward and the negative curvature of a concave topography is upward. That is, according to the WOE and SE methods, the maximum values of (C = 0.57, −0.49 and −0.66) and (Pij = 0.50, 0.29 and 0.21) are related to "Concave," "Convex," and "Flat" classes, respectively. Negative topographies hold more water and retain water due to rainfall for long some cycle of time during rainfall than with positively curved slopes, resulting in increased soil wet [72]. The SFM is the most sensitive component of the slope relative to the landslide in the area studied by "Cherty Quartzit, Dolomite with Epidiorite Dykes", this type of material contributing to the highest number of landslide (67) and also the highest weight landslide (C = 1.45) and density landslide (Pij = 0.35) in accordance with the WOE and SE methods, respectively. The SFM is another unique feature of each region and offers different results in different regions.

Most of the study area belongs to "Sparse vegetation", so it seems quite logical that there are 89 landslides in this land cover. On the other hand, the highest weight (C = 1.59) and density of landslides (Pij = 0.44) based on two algorithms (WOE and SE) is related to "Querry" land cover. The number and weight of landslides in areas covered with vegetation is higher, which is consistent with previous research [73]. Conversely, in the study area on landscaping where there is stagnant water, hard and rocky areas and habitats such as "River," "Barren rocky slope," "Wasteland," and "Settlement" classes, the number of landslides is zero.

Roads have a great impact on landslides [55]. In the present area, with a distance from the roads, the distance of more than 500 meters, the number of landslides has increased a lot, which is due to the expansion of the area, and on the other hand, at distances of 100–200 meters from the road, the amount of landslide weight (C = 0.69) is more and at distances less than 100 meters, the landslide density (Pij = 0.49) is the highest according to the WOE and SE.
The image contains a table titled "Table 2: WOE and SE analysis of landslide conditioning factors." The table is divided into columns for factors, classes, landslide pixels, % class pixels, % landslide pixels, WOE, WSE, C, Pij, and Eij. The table includes data for various factors such as slope, aspect, landforms, curvature, and SFM. The table also includes calculations for landslide pixels and % class pixels, with additional columns for WOE, WSE, C, Pij, and Eij. The data is presented in a tabular format with specific values for each category.
humidity through distance from the river can affect the occurrence of landslides and create a high correlation with the presence of landslides. The results of this variable based on the WOE and its correlation with landslide occurrence are consistent with Pourghasemi, Pradhan, Gokceoglu, Mohammadi, and Moradi [27] study.

The results of the relationship between different classes of elevation and the occurrence of landslides indicate that in the third classes (900–1100 meters) has the highest number (83), weight (C = 1.43) and density ($P_{ij} = 0.33$) of landslides according to both algorithms (WOE and SE). Since in the present study the highest incidence of landslides compared to altitude classes landslide occurrence in the third floor out of 9 floors shows that in the lower floor was the most sensitive to landslides and it can be seen that this factor has little effect on landslides susceptibility and other factors play a greater role in landslides and since the landslide occurs at high altitudes, the different result may be that in the high altitudes of the area the study is due to the rocky nature of the region, which has not occurred on the last floor, the

<table>
<thead>
<tr>
<th>Factors</th>
<th>Class</th>
<th>Class pixels</th>
<th>Landslide pixels</th>
<th>% class pixels</th>
<th>% landslide pixels</th>
<th>WOE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>River</td>
<td>96</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Barren rocky slope</td>
<td>269</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cultivated land</td>
<td>164224</td>
<td>35</td>
<td>22.08</td>
<td>16.99</td>
<td>-0.26</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Extensive slope cut</td>
<td>11006</td>
<td>3</td>
<td>1.48</td>
<td>1.46</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Moderate vegetation</td>
<td>162678</td>
<td>45</td>
<td>21.87</td>
<td>21.84</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Land cover</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plantation</td>
<td>2865</td>
<td>1</td>
<td>0.39</td>
<td>0.49</td>
<td>0.23</td>
<td>0.23</td>
<td>0.11</td>
</tr>
<tr>
<td>Settlement</td>
<td>14667</td>
<td>0</td>
<td>1.97</td>
<td>0</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Quarry</td>
<td>15136</td>
<td>19</td>
<td>2.04</td>
<td>9.22</td>
<td>1.51</td>
<td>-0.08</td>
<td>1.59</td>
</tr>
<tr>
<td>Sparse vegetation</td>
<td>266609</td>
<td>89</td>
<td>35.85</td>
<td>43.2</td>
<td>0.19</td>
<td>-0.12</td>
<td>0.31</td>
</tr>
<tr>
<td>Thick vegetation</td>
<td>104103</td>
<td>14</td>
<td>14.00</td>
<td>6.8</td>
<td>-0.72</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>Wasteland</td>
<td>2053</td>
<td>0</td>
<td>0.28</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

| Distance to roads (m)        |                |              |                  |                |                   |       |      |
| 0–100                        | 67380          | 34           | 9.06             | 16.50          | 0.18               | -0.46 | 0.49 |
| 100–200                      | 57028          | 11           | 7.67             | 5.34           | 0.60               | -0.09 | 0.69 |
| 200–300                      | 49354          | 0            | 6.64             | 0.00           | -0.36              | 0.02  | 0.00 |
| 300–400                      | 42511          | 0            | 5.86             | 0.00           | 0.00               | -0.07 | 0.00 |
| >500                         | 487139         | 161          | 65.50            | 78.16          | 0.00               | -0.05 | 0.32 |
| Overburden depth (m)         |                |              |                  |                |                   |       |      |
| 0–1                          | 480051         | 163          | 64.55            | 79.13          | 0.00               | -0.02 | 0.43 |
| 1–2                          | 203003         | 28           | 27.30            | 35.39          | 0.20               | -0.53 | 0.73 |
| 2–5                          | 48078          | 15           | 6.46             | 7.28           | -0.70              | 0.17  | -0.87 |
| >5                           | 12591          | 0            | 1.69             | 0              | 0.12               | -0.01 | 0.13 |
| Distance to rivers (m)       |                |              |                  |                |                   |       |      |
| 0–100                        | 68250          | 7            | 9.18             | 3.4            | 0.09               | -0.13 | 0.22 |
| 100–200                      | 65780          | 19           | 8.84             | 9.22           | -0.99              | 0.06  | 1.06 |
| 200–300                      | 63677          | 12           | 8.56             | 8.53           | 0.04               | 0.00  | 0.05 |
| 300–400                      | 61359          | 18           | 8.25             | 8.74           | -0.39              | 0.03  | -0.41 |
| >500                         | 58804          | 21           | 7.91             | 10.19          | 0.00               | -0.01 | 0.01 |
| Elevation (m)                |                |              |                  |                |                   |       |      |
| <700                         | 41938          | 5            | 5.64             | 2.43           | -0.84              | 0.03  | -0.88 |
| 700–900                      | 69721          | 51           | 9.37             | 24.76          | 0.97               | -0.19 | 1.16 |
| 900–1100                     | 103488         | 83           | 13.91            | 40.29          | 1.06               | -0.37 | 1.43 |
| 1100–1300                    | 128838         | 36           | 17.32            | 17.48          | 0.01               | 0.00  | 0.01 |
| 1300–1500                    | 145497         | 16           | 19.56            | 7.77           | -0.92              | 0.14  | -1.06 |
| 1500–1700                    | 127267         | 2            | 17.11            | 0.97           | -2.87              | 0.18  | -3.05 |
| 1700–1900                    | 79239          | 0            | 10.65            | 0              | 0.00               | 0.11  | -0.11 |
| 1900–2100                    | 35169          | 13           | 4.73             | 6.31           | 0.29               | -0.02 | 0.31 |
| 2100–2427                    | 12566          | 1            | 0.69             | 0              | 0.00               | 0.02  | 0.00 |
highest elevation of the landslide [39]. Some researchers used altitude as a controlling factor in the occurrence of landslides [75].

In general, the causes of landslides are many, complex and sometimes unknown. Although the underlying factors influencing landslide occurrence can be observed during field visits, aerial photo interpretation, and satellite imagery. Several geomorphometric factors are involved in the analysis to investigate the effective factors in landslide occurrence [76]. Quantitative measurement of many geomorphometric factors by field visits is difficult and therefore it is difficult to know their relationship to the occurrence of landslide mechanism. Because landslides are among the most devastating natural disasters, many researchers around the world have attempted to assess landslide hazards, identify hazardous areas, and display their spatial distribution by indirect methods [20]. In this study, Google Earth images, DEM data and Geological Survey of India maps have been

Figure 5: Landslide susceptibility map using SE model and WOE model.
used for the development of landslide susceptibility maps using the WOE and SE, which is an effective approach for landslide study in regional scale.

4.2. Development of Landslide Susceptibility Maps. Weights of each classes of the factors generated from the WOE and SE methods (Table 2) were used to generate landslide susceptibility maps of the study area using GIS application (Figure 5). Natural break classification method was used to classify the landslide susceptibility indices into five classes. Figure 6 shows the percentage distribution of landslide pixels in five landslide susceptibility classes “very low, low, moderate, high and very high” based on both SE and WOE algorithms. In the very low susceptibility class, 10.91% of the study area is affected by landslide susceptibility according to the SE and almost negligible according to the WOE model. In the low susceptibility class, 10.91% and 5.45% of the area affected by landslides according to the SE and WOE methods, respectively. On the other hand, in the moderate susceptibility class, 40% and 14.55% of the study area is affected according to the SE and WOE methods, respectively. Also, in the high susceptibility class, 30.91% area and 40% area is affected according to the SE and WOE methods, respectively. Whereas, in the high susceptibility class, 7.27% and 40% area is affected according to the SE and WOE methods, respectively. The analysis indicated the WOE method gave high distribution (40%) of landslides in

![Figure 6: Distribution of classes on landslide susceptibility maps.](image1)

![Figure 7: Validation of the WOE and SE models using ROC curve.](image2)
both high and very high susceptibility class compared with those of the SE method.

In this study, the ROC curve was used to evaluate the bivariate statistical models: SE and WOE (Figure 7). Area under this curve of the SE model is 52.17 (AUCSE) and of the WOE model is 68.75 (AUCWOE) which means weak accuracy of prediction of the SE model and moderate of the WOE model in the development of landslide susceptibility zone maps (Figure 7). The results of the accuracy of the method WOE compared to another method are consistent with other researchers [40, 55, 71]; however, their performance is lower than other ML models using the same dataset such as Naïve Bayes (AUC = 0.873), Multilayer Perceptron neural network classifier (AUC = 0.864), and Alternating Decision Tree (AUC = 0.840). One of the advantages of these two-variable statistical methods is that in these models, data collection and analysis is relatively easy and requires little time to do it.

4.3. Validation of Landslide Susceptibility Models. In this study, the ROC curve was used to evaluate the bivariate statistical models: SE and WOE (Figure 7). Area under this curve of the SE model is 52.17 (AUCSE) and of the WOE model is 68.75 (AUCWOE) which means weak accuracy of prediction of the SE model and moderate of the WOE model in the development of landslide susceptibility zone maps (Figure 7). The results of the accuracy of the method WOE compared to another method are consistent with other researchers [40, 55, 71]. One of the advantages of these two-variable statistical methods is that in these models, data collection and analysis is relatively easy and requires little time to do it.

5. Conclusion

In the present study, performances of the two simple popular bivariate statistical models (SE and WOE) have been evaluated for developing landslide susceptibility maps of Pithoragarh district, Uttarakhand state, India. The AUC ROC results indicated that WOE model (AUCWOE = 68.75) is better than the SE model (AUCSE = 52.17). Even though the AUC values of the models are not high, they are acceptable for landslide susceptibility mapping, as the landslide classes boundaries are not regular in Himalayan region and depends on heterogeneous topographical and geological features. Thus, WOE model having better performance can be used for the identification of landslide susceptible zones which can be used for the land use planning and prevention of landslides in hilly and mountainous areas not only Himalayas but other parts of the world also. Nowadays, ML methods are being applied for model studies in landslide susceptibility mapping. It is proposed to carry out Machine Learning model studies in this area and compare the results with bivariate statistical models for further improvement of performance considering more input parameters.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank the University of Transport Technology, Hanoi, Vietnam, for supporting this research.

References


[14] Y. Achour, A. Boumezeur, R. Hadji, A. Chouabbi, V. Cavaleiro, and E. A. Bendaoud, “Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine,
Advances in Civil Engineering


B. T. Pham, B. Pradhan, D. Tien Bui, I. Prakash, and M. B. Dholakia, “A comparative study of different machine learning methods for landslide susceptibility assessment: a...


