



Enhancing landslide susceptibility prediction using deep learning through optimal activation and loss function selection

Article info

Type of article:

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2026.en.6.5.33-50>

*Corresponding author:

Email address:

tphong1617@gmail.com

Received: 12/11/2025

Received in Revised Form:
29/06/2026

Accepted: 08/07/2026

Tran Huu Anh¹, Tran Van Phong^{2,3*}, Dung Nguyen Duy⁴, Nguyen Thi Bich Hanh⁵

¹Faculty of Technology and Engineering, Thai Binh University, Hung Yen, Vietnam; anhuni1006@gmail.com

²Institute of Earth Sciences, Vietnam Academy of Science and Technology, 68 Huynh Thuc Khang, Hanoi, Vietnam; tphong1617@gmail.com

³Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Hanoi, Vietnam

⁴University of Transport Ho Chi Minh City, 2 Vo Oanh, Thanh My Tay, Ho Chi Minh 700000, Vietnam; dung.nguyen@ut.edu.vn

⁵Geotechnical Engineering and Artificial Intelligence research Group, University of Transport Technology, 54 Trieu Khuc, Hanoi, Vietnam; hanhntb@utt.edu.vn

Abstract: Landslide susceptibility prediction has increasingly shifted from traditional statistical methods toward deep learning approaches capable of modeling complex nonlinear relationships among geo-environmental controlling factors. However, model performance is influenced not only by network architecture but also by the selection of activation and loss functions, which directly affect learning dynamics and generalization capability. This study systematically investigates the influence of activation–loss function combinations within a unified deep learning framework for landslide susceptibility mapping in Sin Ho area, northwestern Vietnam. A multi-source geospatial database integrating topographic, geological, hydrological, land-use, seismic, and rainfall-related factors was used to train and validate multiple model configurations. Model performance was evaluated using a comprehensive set of quantitative indicators, including area under the ROC curve (AUC) and accuracy (ACC). The results indicate that models employing the Softmax activation function consistently achieve superior predictive performance and robustness. Among all tested configurations, the Softmax–MultiLabel (SM–ML) model delivered the most reliable results, attaining a validation AUC of 0.878, and ACC of 83.75, demonstrating strong discrimination ability, stable generalization, and reduced prediction error. The landslide susceptibility map generated using the optimal SM–ML configuration exhibits a clear and progressive increase in landslide density from very low to very high susceptibility classes, confirming meaningful spatial stratification of landslide hazard. These findings highlight that careful selection of activation and loss functions can substantially enhance deep learning performance without increasing model complexity or computational cost, providing a robust and efficient framework for landslide susceptibility assessment and

Abstract: (continued) disaster-risk reduction in mountainous regions under changing climatic conditions.

Keywords: Deep learning, Landslide susceptibility, Activation function, Loss function, GIS.

1. Introduction

Landslides represent one of the most destructive hydro-geomorphic hazards, responsible for significant loss of life, infrastructure damage, and socio-economic disruption in mountainous regions worldwide [1-3]. Their occurrence is governed by the interplay of intrinsic conditioning factors such as lithology, slope morphology, fault distribution, land use, and soil properties, and extrinsic triggers, particularly extreme rainfall and anthropogenic disturbances [4-7]. With the increasing frequency and intensity of rainfall events under climate change, reliable spatial prediction of landslide-prone areas has become critical for disaster mitigation, land-use planning, and infrastructure development [8, 9].

Early models for landslide susceptibility assessment employed statistical and heuristic approaches, including frequency ratio, weight-of-evidence, analytical hierarchy process, and logistic regression [10-12]. Although effective for regional screening, these methods generally assume linear relationships among conditioning variables and are limited in reproducing the nonlinear complexity of geomorphological systems. Machine learning and deep learning models such as random forest, support vector machine, convolutional neural networks (CNN), multilayer perceptrons (MLP), and long short-term memory (LSTM) have since demonstrated improved predictive performance because of their ability to capture nonlinear and high-dimensional dependencies among geospatial factors [7, 13-18].

However, most studies focus primarily on network architecture, while activation and loss functions, fundamental drivers of neural learning, receive limited attention despite their strong influence on gradient propagation, class

separability, and probability calibration [19, 20]. In susceptibility mapping, where datasets are highly imbalanced and landslide pixels represent the minority class, suboptimal loss-activation pairing can severely degrade prediction reliability by generating either excessive false stability or excessive false alarms [21, 22]. Therefore, systematic investigation of activation-loss interactions is critical for developing robust deep learning frameworks for geohazard prediction [23]. To address this gap, the present study evaluates nine loss functions and two activation functions under a unified neural architecture using a high-quality landslide inventory and multi-criteria geospatial database from the Sin Ho area, Northwest Vietnam, an area characterized by steep terrain, complex geology, and frequent rainfall-triggered failures [24]. By isolating the effect of activation-loss pairing while holding all other design elements constant, this work provides direct evidence on the contribution of learning objectives to susceptibility performance. The aims of the study are to: (1) quantify the effect of activation-loss combinations on classification accuracy, sensitivity, specificity, AUC, and generalization; (2) identify the optimal configuration for spatial prediction of landslide susceptibility; and (3) evaluate the geomorphological validity of susceptibility patterns produced by the best model. A landslide-prone Sin Ho area of Lai Chau Province was selected as a case study for generation of the spatial database for modeling. Various statistical metrics, including the ROC curve, were used for validation and comparison of the models.

2. Study area

The study was conducted in the highly landslide-prone Sin Ho area of Lai Chau Province

in northwestern Vietnam [25]. The Sin Ho area covers an area of approximately 1,526.96 km². Geographically, it extends from about 22°02' to 22°37' north latitude and from 102°56' to 103°24' east longitude (Fig. 1). The terrain is predominantly mountainous, characterized by strongly dissected hills with alternating valleys and ridgelines that generally follow a northwest–southeast orientation consistent with regional geological structures. The geological setting is highly complex, with intense tectonic deformation resulting in elevated relief and steep slopes. Altitudes across the district range from roughly 92 m to nearly 2,500 m above sea level. Surface materials are mainly composed of weathered rock formations overlain by soil layers and vegetation cover. In addition, the district is traversed by a dense and well-developed network of rivers and streams.

The climate of the Sin Ho area reflects typical tropical conditions of the northwestern mountainous region and is strongly influenced by monsoonal weather systems. Two distinct seasons are observed: a rainy season and a dry season. Mean annual precipitation is approximately 2,604 mm, with rainfall distribution being highly uneven throughout the year. Upland communes generally receive between 2,600 and 2,700 mm of rainfall annually, while lowland areas and communes located along the Nam Na River record rainfall ranging from about 2,480 to 2,750 mm per year. Peak precipitation usually occurs in June and July, accounting for nearly 70% of the total annual rainfall. Average relative humidity varies from 80% to 86% annually, reaching its maximum in July (around 85–90%) and dropping to its lowest levels in March, typically between 70% and 80%.

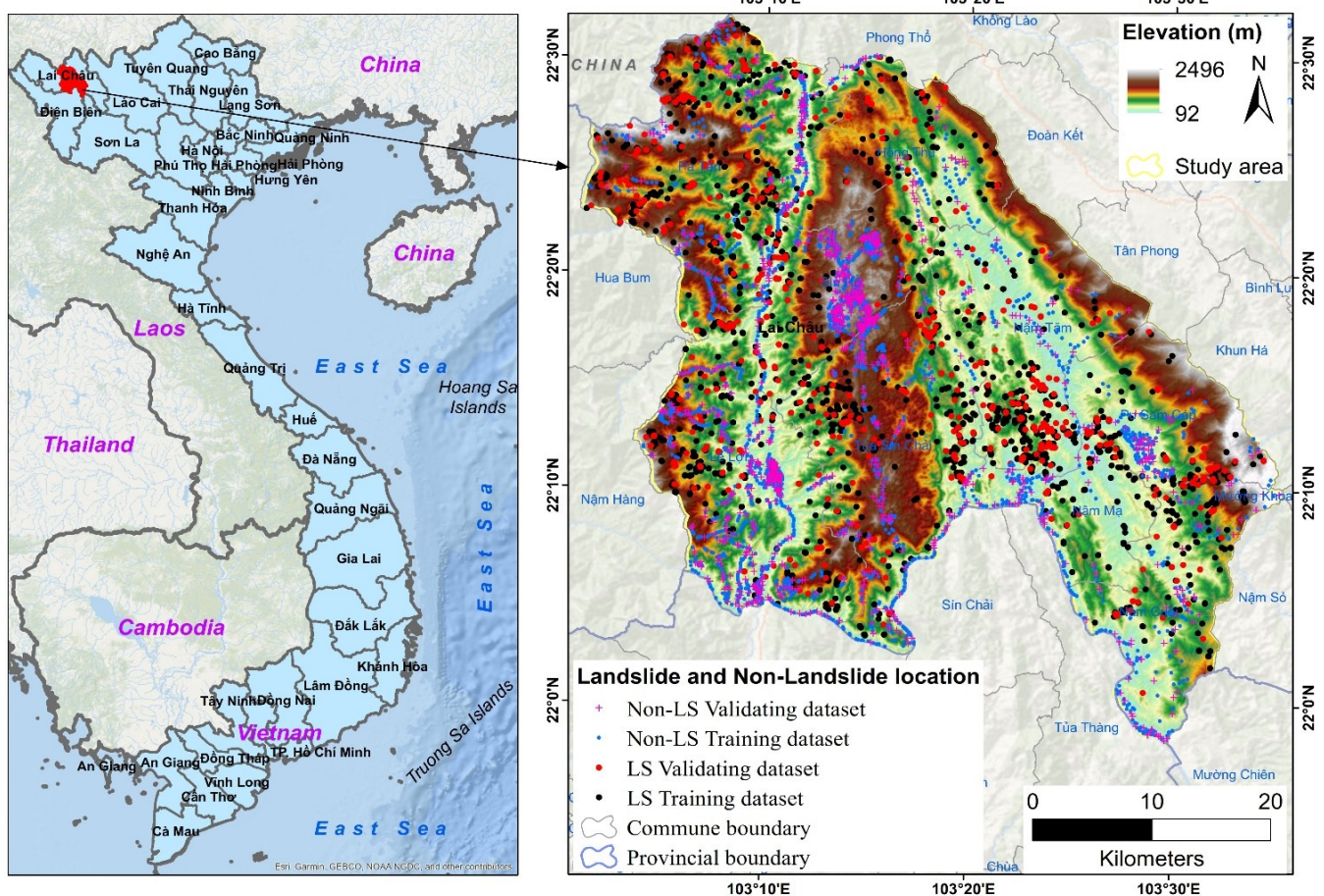


Fig. 1. The study area of Sin Ho, Lai Chau province, Vietnam

3. Methods and data used

3.1. Methodological flow chart

The methodological flow chart of this study was structured into six sequential stages to ensure

a systematic and reliable analysis (Fig. 2). First, the study area was selected based on geomorphological characteristics and the availability of relevant data. Next, multi-source datasets were collected to support model development and validation. These datasets were then processed, filtered, and normalized to ensure consistency and suitability for machine learning applications. In the modeling phase, different

combinations of activation functions and loss functions were tested to identify the most effective configuration for landslide prediction. The resulting models were evaluated using multiple performance metrics to determine predictive accuracy and generalization capability. Finally, the best-performing model was applied to generate a landslide susceptibility map for the study area, providing a spatial representation of risk zones.

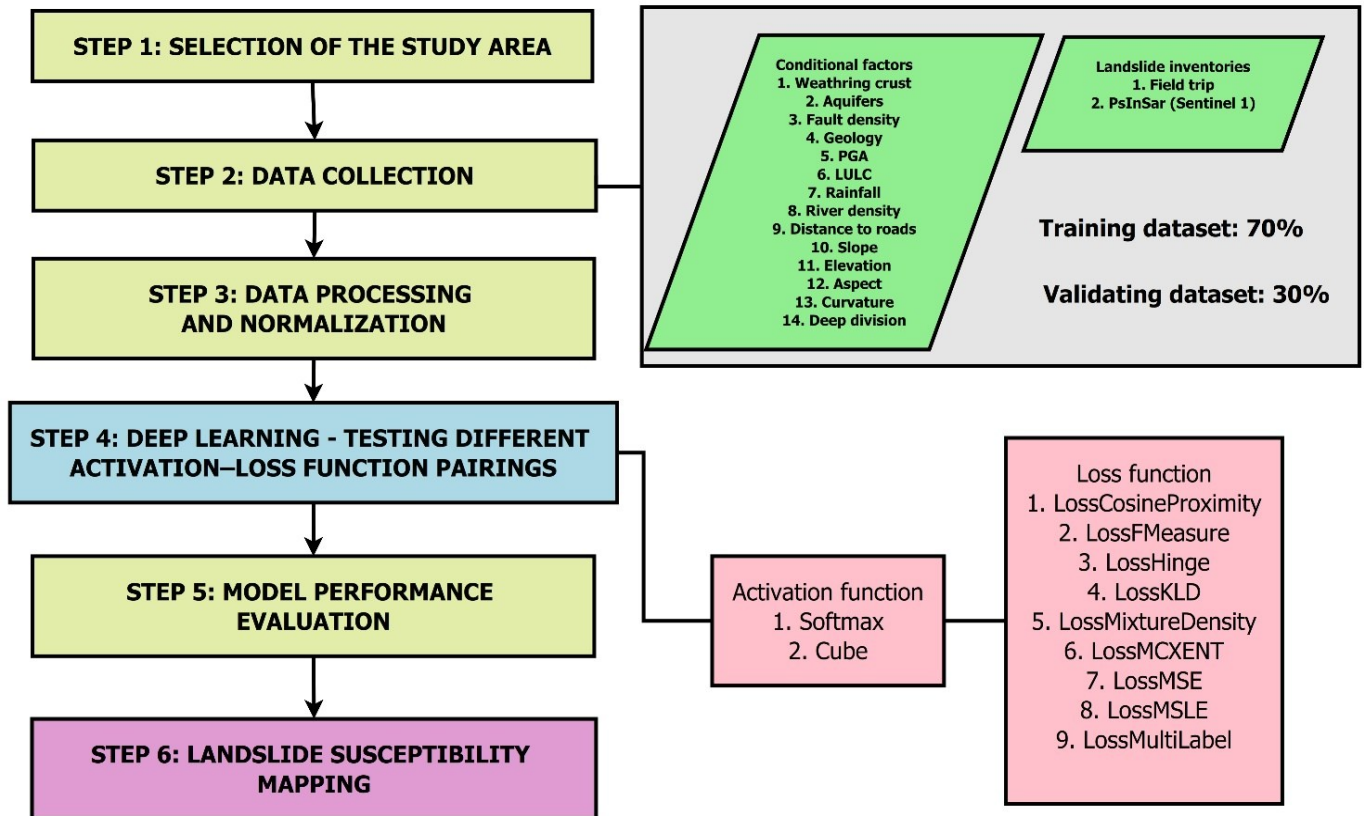


Fig. 2. Methodology flow of this research

3.2. Methods used

3.2.1. Deep Learning

Deep learning is a subfield of machine learning that focuses on algorithms inspired by the structure and functioning of the human brain [26]. It is based on artificial neural networks with multiple hidden layers, which enable the model to automatically learn complex patterns and representations from large volumes of data. By progressively extracting higher-level features, deep learning models are particularly effective in handling nonlinear relationships and high-dimensional data. In recent years, deep learning

has achieved remarkable success in various domains such as image recognition, natural language processing, speech recognition, and predictive modeling, making it a powerful tool for solving complex real-world problems, including natural hazard modeling and assessment [27]. However, the performance of deep learning models depends significantly on the selection and optimization of the activation function and loss function used in training the models [28].

3.2.1.1. Activation Function

Activation functions introduce non-linearity into neural networks, enabling them to learn

complex patterns that cannot be captured by purely linear transformations [29]. Without them, even deep architectures collapse into a single linear mapping, limiting their capacity to model non-linear relationships. Therefore, selecting appropriate activation functions is essential for effective deep learning model design.

In this study, we employ two activation functions: Softmax and Cube. Softmax is used in the output layer for multi-class classification, converting raw logits into a normalized probability distribution across classes [30]. The Cube function, applied within hidden layers, raises its input to the third power, providing a non-linear transformation that supports feature abstraction and hierarchical pattern learning [31].

Mathematically, for an input vector of scores $z = [z_1, z_2, \dots, z_k]$, the Softmax function $\sigma(z)$ produces an output vector $p = [p_1, p_2, \dots, p_k]$ where each component is given by the formula:

$$\sigma(z)_i = \frac{\exp(z_i)}{\sum_{j=1}^K \exp(z_j)} \quad (1)$$

In Equation (1), $\exp(z_i)$ denotes the exponential of z_i . This formula ensures that each transformed score p_i is in the range (0,1) and that

all outputs sum up to 1, (i.e., $\sum_{i=1}^K p_i = 1$). Intuitively,

the exponentials make any one score's contribution grow exponentially with its magnitude, and dividing by the sum of all exponentials means each output can be seen as the fraction of the total "exp weight" contributed by that class. If one particular score z_k is much larger than the others, e^{z_k} will dominate the sum and p_k will be close to 1 (with others near 0), indicating a clear prediction for class k .

The Cube activation function is a non-linear function that raises its input value to the third power. In other words, given an input x (which typically would be the summed input to a neuron,

i.e., $w^T x + b$ for some weight vector and bias), the Cube activation produces x^3 as the output. This can be written succinctly as:

$$g(x) = x^3 \quad (2)$$

Equation (2) defines the Cube activation function $g(x)$. This function is an odd polynomial transformation: it is monotonic and retains the sign of the input (since cubing a negative number gives a negative result and cubing a positive yield positive result). Unlike more common activations such as the sigmoid or hyperbolic tangent, the cube function does not squash its outputs into a fixed range – its output can range from $-\infty$ to $+\infty$, growing in magnitude cubically as the input grows. This means that for larger input values, the Cube activation will produce very large outputs (and similarly very large gradients in magnitude, since $g'(x) = 3x^2$, whereas inputs with small magnitude (between -1 and 1) will become even smaller when cubed (e.g. $0.5^3 = 0.125$). These properties make the cube function quite different from standard rectified linear or sigmoid-style activations, offering a stronger non-linearity that can amplify differences between inputs.

In this study, the Cube activation function is leveraged in a hidden layer to allow the model to capture higher-order interactions among features. The key intuition is that by cubing the linear combination of inputs, the network effectively introduces multiplicative interactions between different input components. For example, if a hidden neuron receives a weighted sum $s = w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + b$ from various input features, its activated output will be s^3 . When expanded, s^3 contains terms proportional to products like $w_1 w_2 w_3$ (and other combinations of input variables up to third degree). In other words, that neuron's output inherently includes terms that represent the three-way interaction between x_1, x_2, x_3 (scaled by the product of the

corresponding weights), as well as squared terms and linear terms. This ability to capture third-order feature interactions directly within one activation is a potential advantage in tasks where the prediction depends on a combination of three different factors.

3.2.2.2. Loss Function

Loss functions are a critical component in training deep learning models, as they quantify the error between a model's predictions and the true target values [32]. During training, optimization algorithms adjust model parameters to minimize the loss, thereby improving the model's performance. Different loss functions are used depending on the task and the nature of the output (e.g., classification probabilities, regression values, etc.). Below, we introduce several loss functions LossCosineProximity, LossFMeasure, LossHinge, LossKLD, LossMixtureDensity, LossMCXENT, LossMSE, LossMSLE, and LossMultiLabel – along with their mathematical formulations and typical use cases in deep learning [32].

Loss functions guide the training of deep learning models by defining an objective to minimize [33]. The choice of loss function depends on the problem: LossCosineProximity for vector similarity tasks, LossFMeasure for directly optimizing classification metrics, LossHinge for margin-based classification, LossKLD for matching probability distributions, LossMixtureDensity for modeling complex output distributions, LossMCXENT for standard multi-class classification, LossMSE for regression, LossMSLE for regression with a focus on relative error, and LossMultiLabel for problems with multiple simultaneous labels. Each provides a different way of quantifying error, and understanding their formulations helps in selecting the appropriate loss for a given deep learning application.

3.2.2. Validation metrics

In this study, several widely used statistical indicators including Sensitivity, Specificity, Positive

Predictive Value (PPV), Negative Predictive Value (NPV), Accuracy (ACC), Root Mean Square Error (RMSE), and the Area Under the Receiver Operating Characteristic curve (AUC–ROC) [9, 24] were employed to validate and compare the predictive capability of the developed models. These indicators have been extensively adopted in previous landslide susceptibility studies and provide a comprehensive assessment of classification accuracy and error characteristics.

To ensure an objective evaluation, model performance was examined separately for the training and validation datasets, as recommended for landslide prediction analysis. In general, higher values of Sensitivity, Specificity, PPV, NPV, ACC, and AUC indicate stronger predictive performance, whereas lower RMSE values reflect reduced prediction error and improved model reliability. Together, these complementary metrics allow for a robust and balanced comparison of model accuracy, discrimination ability, and error behavior.

3.3. Data used

3.3.1. Landslide inventory

The landslide inventory used in this study was developed using the Persistent Scatterer Interferometric Synthetic Aperture Radar (PSInSAR) technique, employing multi-temporal Sentinel-1 SAR imagery provided by the European Space Agency (ESA) [34]. The dataset was updated and refined from the previous work by Anh et al. (2024) [24]. Persistent Scatterers (PS) were extracted from radar image stacks to produce ground displacement time series with millimeter-level accuracy. Areas showing continuous surface deformation beyond a set threshold were classified as active or potential landslides, creating a comprehensive inventory for the study area.

The integration of Sentinel-1 data in PSInSAR offers substantial advantages due to its ability to acquire imagery regardless of cloud cover, illumination, or weather conditions, along with wide-area coverage and high temporal frequency [35]. These characteristics are

particularly valuable in mountainous regions with limited field accessibility and frequent optical obstruction. The resulting inventory provides both spatial landslide distribution and deformation velocity, serving as a fundamental dataset for correlation analysis with conditioning factors, model calibration, and susceptibility prediction. This inventory constitutes a core source of information, enhancing the reliability and applicability of landslide hazard assessment within the region.

In this study, landslide locations were identified and standardized through an integrated

approach combining previously published inventories [24, 25], additional interpretation of radar-based remote sensing imagery, and field verification. As a result, a total of 3016 landslide points were delineated within the study area to construct the landslide inventory database. For model development, the inventory was randomly divided into two independent subsets, with 70% of the polygon data (1406 polygons, 2,090 landslide points) used for training the susceptibility models and the remaining 30% (603 polygons, 926 points) employed for model validation and performance evaluation.

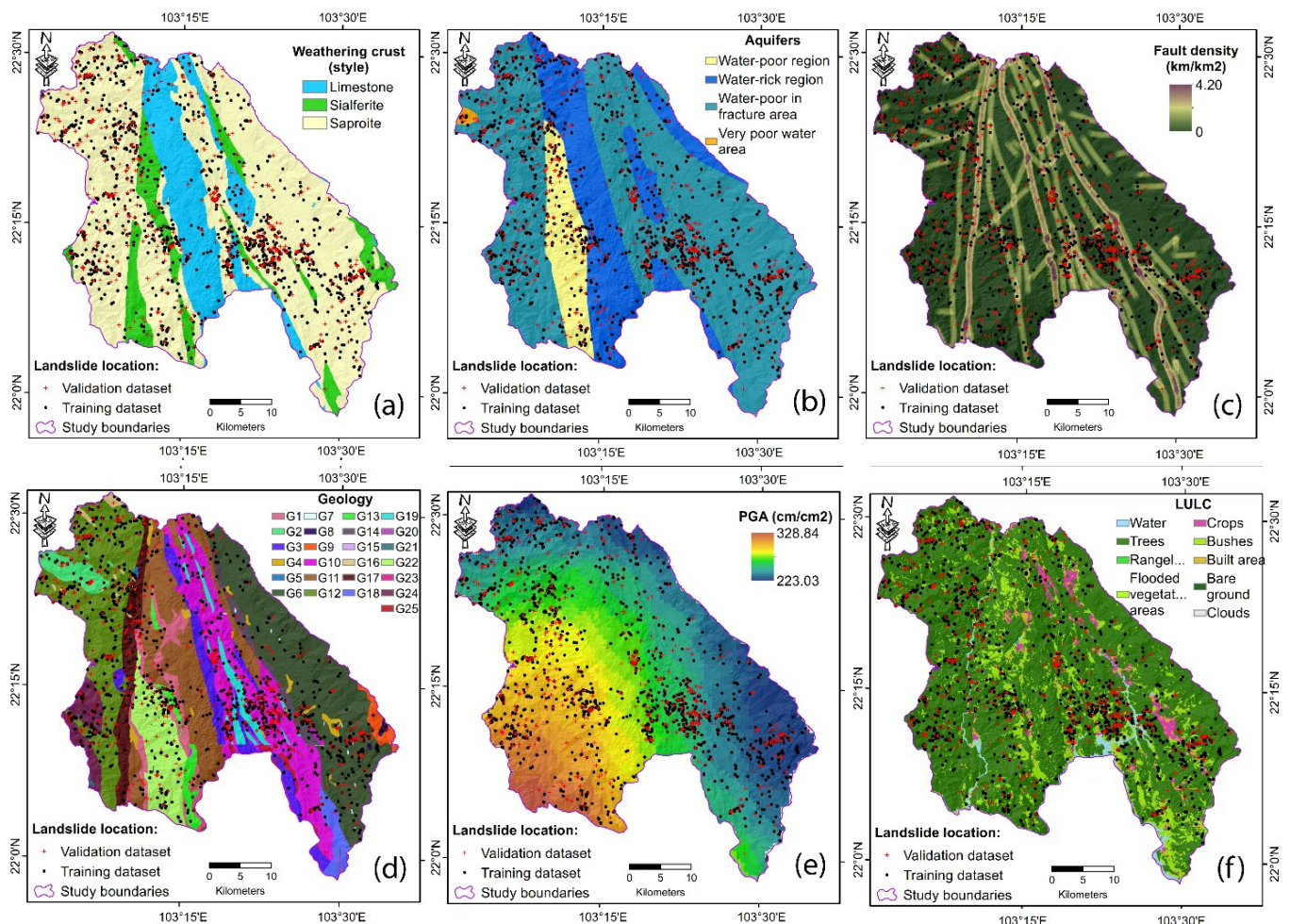


Fig. 3. Conditional factor maps in Sin Ho, Lai Chau

3.3.2. Landslide conditional factors

In this study, fourteen landslide conditioning factors were selected to generate thematic maps for landslide susceptibility analysis. These factors include weathering crust, aquifer, fault density, geology, peak ground acceleration (PGA), land

use/land cover (LULC), rainfall (rainy season), river density, distance to roads, slope, elevation, aspect, curvature, and deep division (Fig. 3). The selection of these parameters was based on local geological–geomorphological conditions, climatic characteristics, seismic setting, human activities,

and previous landslide susceptibility studies, ensuring that the chosen variables adequately

represent both triggering and conditioning mechanisms of landslides.

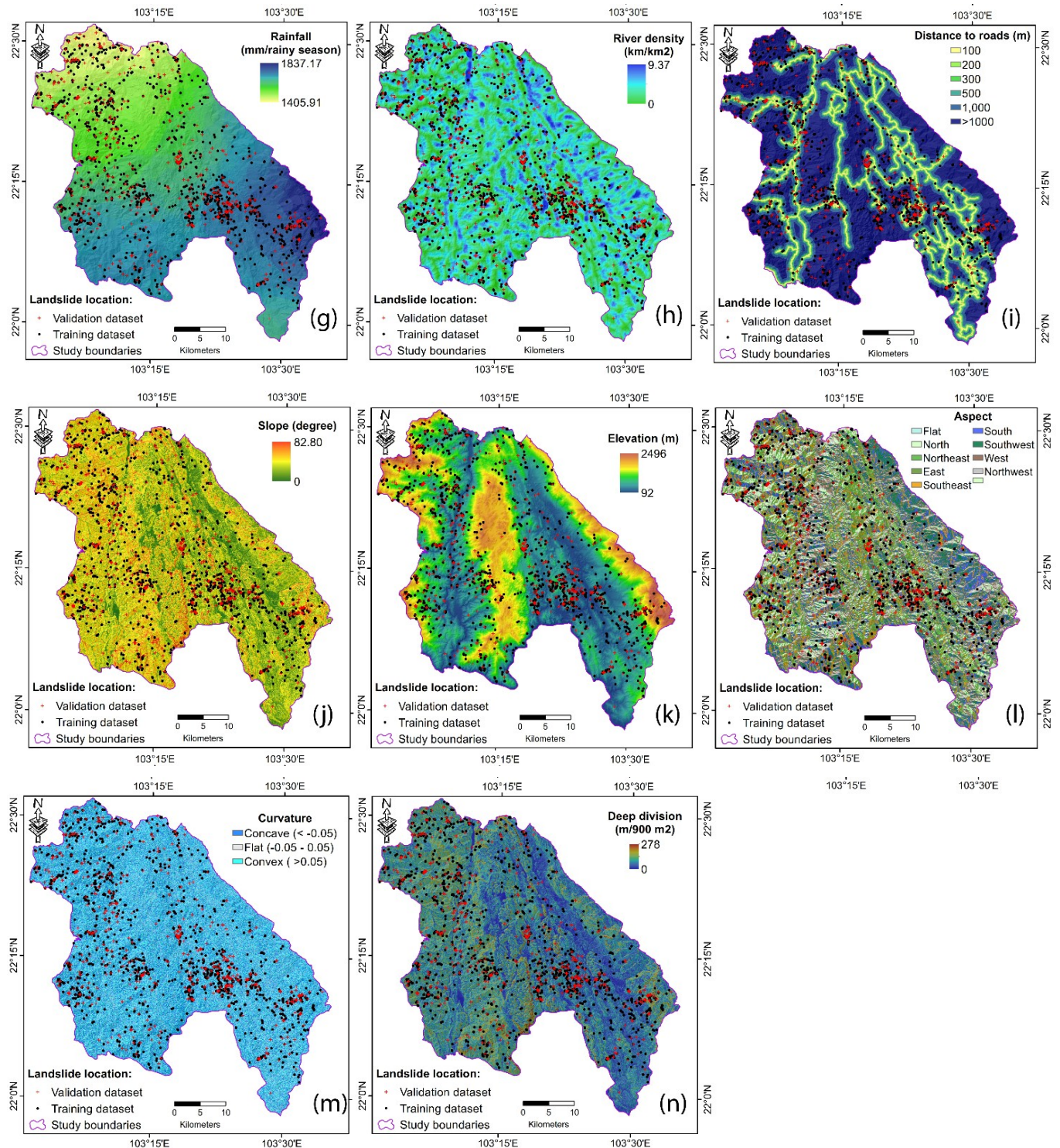


Fig. 3. (continued)

Topographic parameters (slope, elevation, aspect, curvature, deep division, and river density) were derived from a Digital Elevation Model (DEM) with a spatial resolution of 30 m, obtained from the U.S. Geological Survey (USGS). These factors

describe terrain morphology, landscape dissection, and slope geometry, which strongly influence gravitational forces, runoff concentration, and material accumulation on hillslopes.

Geological and hydrogeological parameters,

including weathering crust, geology, fault density, and aquifer, were extracted from 1:200,000-scale geological and hydrogeological maps provided by the Department of Geology and Minerals of Vietnam. These datasets reflect lithological characteristics, tectonic fracturing, groundwater occurrence, and degrees of rock weathering, all of which directly affect slope strength and subsurface hydrological conditions.

Climatic and seismic factors, namely rainfall during the rainy season and peak ground acceleration (PGA), were incorporated to represent major landslide triggers in the study area. Rainfall data were obtained from regional meteorological records and interpolated spatially, while PGA data were derived from seismic hazard maps reflecting regional tectonic activity

Human-induced factors, including land use/land cover (LULC) and distance to roads, were generated from remote sensing imagery and existing infrastructure datasets. These parameters capture the influence of vegetation removal, slope modification, and road construction on slope

instability. All fourteen factors were standardized into raster format with uniform spatial resolution to construct consistent thematic maps, which were subsequently used as input variables for landslide susceptibility modeling.

4. Results

4.1. Model performance based on activation–loss combinations

The performance of the proposed deep learning models was systematically evaluated using different activation–loss function combinations, based on both training and validation datasets. The assessment employed multiple quantitative metrics, including AUC, ACC, Kappa, RMSE, PPV, NPV, SST, and SPF, allowing a comprehensive comparison of predictive accuracy, robustness, and error behavior.

The ROC curves (Figs. 4a–b) reveal clear performance differences among the tested activation–loss configurations. Overall, models using the Softmax activation function consistently outperformed those employing the Cube activation, particularly in terms of AUC values.

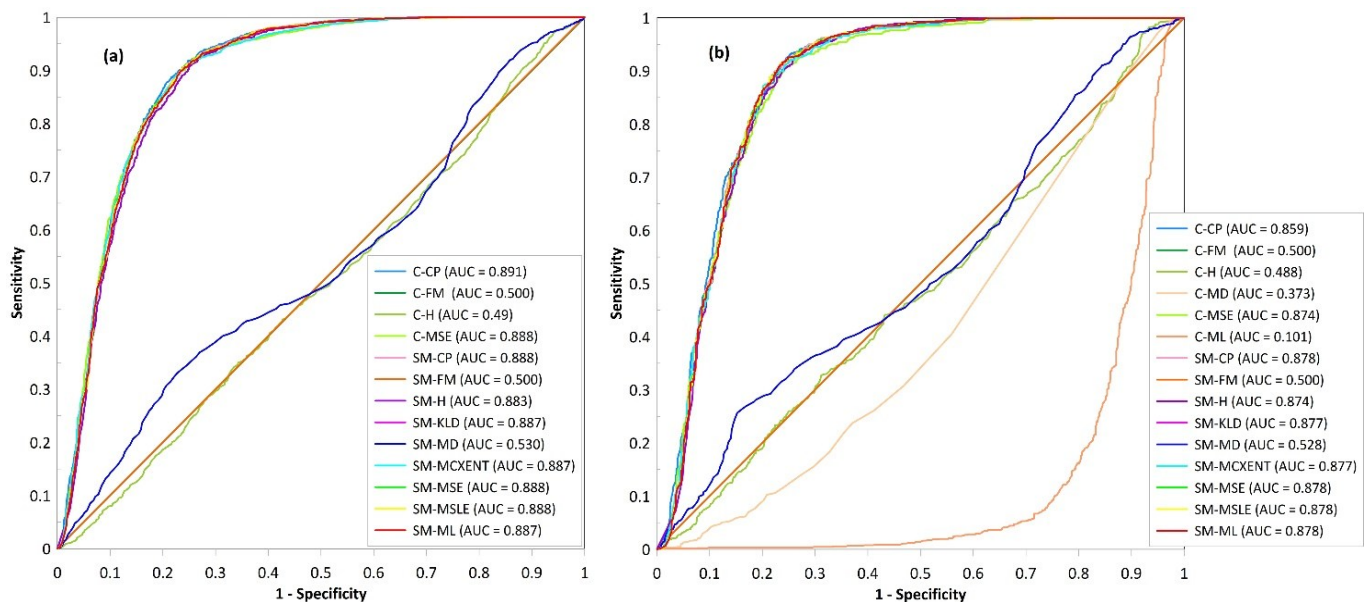


Fig. 4. AUC curves: (a) Training dataset (b) Validating dataset

For the training dataset, the highest AUC values (≈ 0.887 – 0.891) were obtained, notably by Cube–LossCosineProximity (AUC = 0.891) and Softmax–LossMSE / LossMSLE / LossMultiLabel (AUC ≈ 0.887 – 0.888). These curves approach the

upper-left corner of the ROC space, indicating excellent discrimination between landslide and non-landslide classes.

In contrast, some activation–loss combinations exhibited near-random or poor

discrimination capability. For instance, Softmax–LossMeasure and Softmax–LossFM yielded AUC values close to 0.50, while Cube–LossMultiLabel

showed extremely poor performance (AUC $\approx$ 0.10 in validation), representing the weakest model among all tested configurations.

Table 1. The evaluation metrics PPV, NPV, SST, and SPF of the training models

Activation function	Loss function	Training model							
		TP	TN	FP	FN	PPV (%)	NPV (%)	SST (%)	SPF (%)
Softmax	LossCosineProximity	1586	1886	492	204	76.32	90.24	88.60	79.31
	LossFMeasure	0	2090	2078	0	0.00	100.00	-	50.14
	LossHinge	1559	1891	519	199	75.02	90.48	88.68	78.46
	LossKLD	1621	1838	457	252	78.01	87.94	86.55	80.09
	LossMixtureDensity	0	2090	2078	0	0.00	100.00	-	50.14
	LossMCXENT	1621	1838	457	252	78.01	87.94	86.55	80.09
	LossMSE	1592	1878	486	212	76.61	89.86	88.25	79.44
	LossMSLE	1594	1876	484	214	76.71	89.76	88.16	79.49
	LossMultiLabel	1595	1874	483	216	76.76	89.67	88.07	79.51
Cube	LossCosineProximity	114	42	97	3	54.03	93.33	97.44	30.22
	LossFMeasure	0	2090	2090	0	0.00	100.00	-	50.00
	LossHinge	2015	29	63	2061	96.97	1.39	49.44	31.52
	LossKLD	-	-	-	-	-	-	-	-
	LossMixtureDensity	-	-	-	-	-	-	-	-
	LossMCXENT	-	-	-	-	-	-	-	-
	LossMSE	805	1164	320	121	71.56	90.58	86.93	78.44
	LossMSLE	-	-	-	-	-	-	-	-
	LossMultiLabel	-	-	-	-	-	-	-	-

As summarized in Table 3, Softmax-based models achieved superior training accuracy and agreement. Except for two models, Softmax–LossFMeasure and Softmax–LossMixtureDensity, which exhibited poor performance, other models with the Softmax activation function demonstrated strong learning capability and stable convergence. The performance of these models all reached AUC values of approximately 0.888, ACC of approximately 83%, and RMSE around 0.36.

Regarding classification balance, Table 1 shows that Softmax–CosineProximity and Softmax–LossMultiLabel achieved relatively high PPV ($\approx$ 76%), NPV ($>$ 89%), and SST ($>$ 86%), confirming their ability to correctly identify both landslide and non-landslide cases.

Conversely, Cube-based models generally underperformed during training. Several Cube–

loss combinations failed to produce meaningful results or showed very low SST and ACC values. For example, Cube–LossFM and Cube–LossMultiLabel exhibited unstable learning behavior, reflected by low ACC ($<$ 50%) and unreliable Kappa values.

Validation results (Tables 2 and 3) confirm the superiority of Softmax-based configurations in terms of generalization. The Softmax–LossMultiLabel model remained the best-performing model, achieving AUC = 0.878, ACC = 83.75%, Kappa = 0.675, and RMSE = 0.365, indicating a good balance between accuracy and error minimization. Similarly, Softmax–LossCosineProximity, Softmax–LossCXENT and Softmax–LossMSE produced competitive validation AUC values ($\sim$ 0.878) with ACC values above 83%.

Table 2. The evaluation metrics PPV, NPV, SST, and SPF of the validating models

Activation function	Loss function	Validating model							
		TP	TN	FP	FN	PPV (%)	NPV (%)	SST (%)	SPF (%)
Softmax	LossCosineProximity	692	855	233	72	74.81	92.23	90.58	78.58
	LossFMeasure	0	927	925	0	0.00	100.00	-	50.05
	LossHinge	683	856	242	71	73.84	92.34	90.58	77.96
	LossKLD	714	831	211	96	77.19	89.64	88.15	79.75
	LossMixtureDensity	0	927	925	0	0.00	100.00	-	50.05
	LossMCXENT	714	831	211	96	77.19	89.64	88.15	79.75
	LossMSE	698	851	227	76	75.46	91.80	90.18	78.94
	LossMSLE	698	851	227	76	75.46	91.80	90.18	78.94
	LossMultiLabel	701	850	224	77	75.78	91.69	90.10	79.14
Cube	LossCosineProximity	476	663	193	62	71.15	91.45	88.48	77.45
	LossFMeasure	0	927	925	0	0.00	100.00	-	50.05
	LossHinge	896	19	29	908	96.86	2.05	49.67	39.58
	LossKLD	-	-	-	-	-	-	-	-
	LossMixtureDensity	375	0	0	394	100.00	0.00	48.76	-
	LossMCXENT	-	-	-	-	-	-	-	-
	LossMSE	659	797	222	78	74.80	91.09	89.42	78.21
	LossMSLE	-	-	-	-	-	-	-	-
	LossMultiLabel	99	63	359	547	21.62	10.33	15.33	14.93

Table 3. The evaluation metrics AUC, ACC, Kappa, and RMSE of the models

Activation function	Loss function	Time training (s)	Training				Validating			
			AUC	ACC	Kappa	RMSE	AUC	ACC	Kappa	RMSE
Softmax	LossCosineProximity	43.95	0.888	83.30	0.665	0.364	0.878	83.53	0.670	0.369
	LossFMeasure	39.57	0.500	50.14	0.498	0.706	0.500	50.05	0.499	0.706
	LossHinge	32.99	0.883	82.77	0.655	0.368	0.874	83.09	0.661	0.369
	LossKLD	31.12	0.887	82.98	0.659	0.365	0.877	83.42	0.668	0.371
	LossMixtureDensity	0.28	0.530	50.14	0.496	0.568	0.528	50.05	0	0.567
	LossMCXENT	23.22	0.887	82.98	0.659	0.365	0.877	83.42	0.668	0.371
	LossMSE	30.19	0.888	83.25	0.665	0.362	0.878	83.64	0.673	0.368
	LossMSLE	33.24	0.888	83.25	0.665	0.362	0.878	83.64	0.673	0.367
	LossMultiLabel	70.62	0.887	83.23	0.665	0.360	0.878	83.75	0.675	0.365
Cube	LossCosineProximity	37.8	0.891	60.93	0.26	0.536	0.859	81.70	0.630	0.380
	LossFMeasure	36.16	0.500	50.14	0.498	0.706	0.500	50.05	0.499	0.706
	LossHinge	26.04	0.490	49.04	-0.016	0.532	0.488	49.40	-0.010	0.533
	LossKLD	-	-	-	-	-	-	-	-	-
	LossMixtureDensity	-	-	-	-	-	0.373	48.76	0.516	0.713
	LossMCXENT	-	-	-	-	-	-	-	-	-
	LossMSE	23.41	0.888	81.70	0.628	0.391	0.874	82.91	0.658	0.380
	LossMSLE	-	-	-	-	-	-	-	-	-
	LossMultiLabel	-	-	-	-	-	0.101	15.16	-0.647	0.716

In terms of detection capability, Softmax–LossMultiLabel maintained high SST (90.1%) and SPF (79.14%), demonstrating strong sensitivity to landslide occurrences while preserving acceptable specificity.

On the other hand, Cube-based models exhibited clear weaknesses during validation. The Cube–LossMultiLabel model performed worst, with $AUC = 0.101$, $ACC = 15.16\%$, $Kappa = -0.647$, and $RMSE = 0.716$, indicating severe misclassification and poor predictive reliability. This confirms that Cube activation is unsuitable for the studied landslide susceptibility task.

Based on the combined analysis of ROC curves and evaluation metrics from training and validation datasets, Softmax–LossMultiLabel can be identified as the best-performing model, offering the highest discrimination ability, stable accuracy, and low prediction error. In contrast, Cube–LossMultiLabel represents the worst-performing configuration, characterized by extremely low AUC, poor accuracy, and negative agreement, rendering it unsuitable for practical landslide susceptibility modeling.

Overall, the results clearly demonstrate that model performance is highly sensitive to the choice of activation and loss functions, and that Softmax-based deep learning models, particularly when coupled with MultiLabel, Cross-Entropy, or MSE-type loss functions, provide the most reliable and robust predictions for landslide susceptibility assessment.

4.2. Landslide susceptibility mapping

As the SM–ML model exhibited the highest predictive performance among all tested configurations, its outputs were selected to generate the final landslide susceptibility map for the study area (Fig. 5). The resulting susceptibility map was classified into five susceptibility levels, namely very low, low, moderate, high, and very high, using the natural breaks classification method. This zonation provides a clear spatial differentiation of landslide-prone and relatively

stable areas across Sin Ho District.

The spatial reliability of the SM–ML -based susceptibility map was quantitatively evaluated using the frequency ratio (FR) method (Table 4). The results indicate a strong and consistent correspondence between susceptibility classes and observed landslide occurrences. Only about 7% of documented landslides are located within the Very Low and Low susceptibility zones, which together account for approximately 24.3% of the total area. In contrast, nearly 76% of landslides are concentrated in the High and Very High susceptibility zones, despite these classes covering about 59.8% of the study area, highlighting the model's strong discriminative capability. Furthermore, the FR values for landslide pixels (FR_{LS}) increase monotonically from 0.13 in the Very Low class to 1.25 in the *very high* class, indicating a progressive rise in landslide probability with increasing susceptibility level. Conversely, the FR values for non-landslide pixels (FR_{Non-LS}) decrease sharply from 5.03 (very low) to 0.25 (very high), confirming effective separation between stable and unstable terrain conditions. This inverse trend between FR_{LS} and FR_{Non-LS} clearly demonstrates the robustness of the SM–ML model in distinguishing landslide-prone zones.

From a geomorphological perspective, the high and very high susceptibility zones are predominantly distributed over steep, rugged terrain characterized by deep valley incision, high slope gradients, dense drainage networks, and structurally weakened geological formations. These areas are commonly associated with intense weathering, active fluvial erosion, and anthropogenic disturbances such as road cuts. In contrast, the very low and Low susceptibility zones are mainly located on gentle slopes, valley floors, and relatively stable upland surfaces, where landslide occurrences are sparse.

The observed landslide inventory shows a pronounced clustering within the SM–ML-predicted high and very high susceptibility areas, with limited

landslide occurrences in the lower susceptibility classes. This spatial agreement indicates that the proposed model achieves high detection capability with low false-negative and false-positive rates, further confirming its suitability for landslide

susceptibility mapping. Overall, the SM–ML-derived susceptibility map provides a reliable and practical basis for land-use planning, infrastructure development, and landslide risk mitigation in mountainous regions.

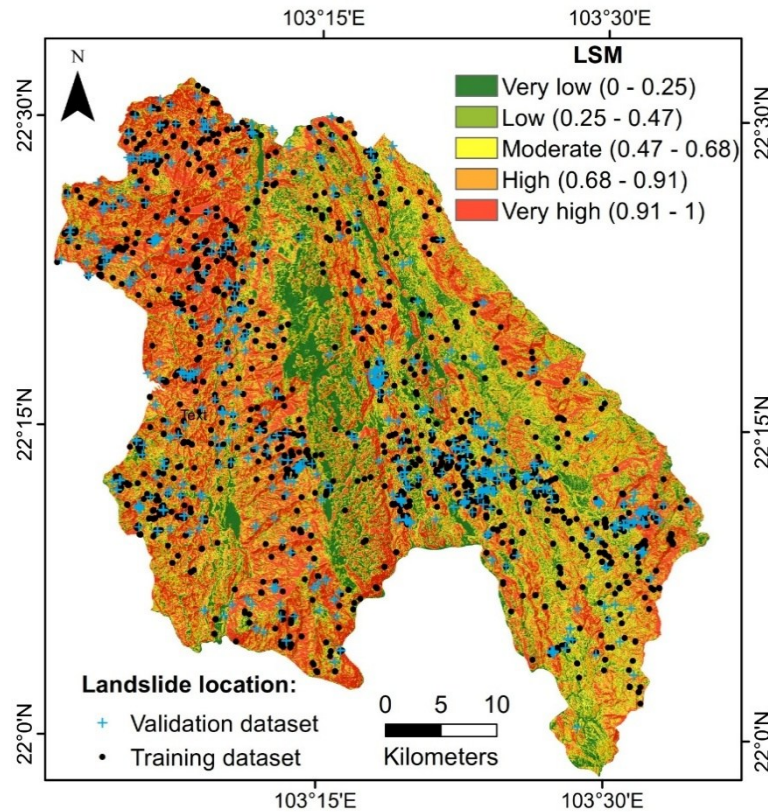


Fig. 5. Landslide susceptibility mapping generated from the SM–ML model

Table 4. Analysis of landslide susceptibility map generated by the best performance model

Landslide Susceptibility Class	LS class pixels	% LSM class	LS pixel	Non-LS pixel	% LS	% Non-LS	FR LS	FR Non-LS
Very low	237177	11.17	13	519	1.40	56.17	0.13	5.03
Low	278857	13.13	53	159	5.72	17.21	0.44	1.31
Moderate	337589	15.90	158	94	17.06	10.17	1.07	0.64
High	636763	29.98	356	84	38.44	9.09	1.28	0.30
Very High	633454	29.83	346	68	37.37	7.36	1.25	0.25

5. Discussion and Conclusion

This study demonstrates that the selection of activation–loss function combinations plays a decisive role in determining the predictive performance of deep learning models for landslide susceptibility mapping, and that this design choice can be as influential as network architecture itself. Through a systematic evaluation of eighteen activation–loss configurations, the results clearly

show that Softmax-based models consistently outperform Cube-based counterparts, owing to their probabilistic output structure and stable gradient behavior [36, 37]. In contrast, Cube activation produces unbounded scores that interact poorly with most classification-oriented loss functions, often leading to unstable learning and poor generalization.

Within the Softmax family, the SM–ML

configuration emerged as the most effective and balanced model. It achieved high validation AUC, accuracy, and Cohen's Kappa with only a small generalization gap between training and validation datasets, while also producing a geomorphologically coherent landslide susceptibility map. The multi-label loss appears particularly well suited to heterogeneous mountainous environments, where conditioning factors influence stable and unstable terrain in partially overlapping ways. By contrast, conventional cross-entropy losses (e.g., MCXENT) remain strong and reliable baselines, but the slightly superior behavior of SM–ML confirms that even modest modifications to the learning objective can yield measurable performance gains without increasing model depth or complexity.

Compared with many recent studies that emphasize increasingly complex architectures such as deep CNNs, hybrid CNN–LSTM models, or attention-based networks, this work shows that optimizing the learning objective alone can elevate model performance into the upper range of reported results. Whereas regional-scale landslide susceptibility studies typically report validation AUC values around 0.80–0.85, the optimized SM–ML configuration achieves higher discrimination and accuracy, while remaining computationally efficient and easier to deploy. This is particularly important for operational applications, where simpler but well-tuned models are more transparent, transferable, and maintainable for local authorities than highly specialized architectures. The ROC curves further support this conclusion, as SM–ML exhibits steep and concave validation curves, indicating high sensitivity at relatively low false-positive rates.

The results also clarify important operational trade-offs among different loss-function families. Regression-type losses (e.g., MSE and MSLE) combined with Softmax tend to favor the non-landslide class, yielding high specificity but inadequate sensitivity. Such configurations risk

missing a substantial proportion of landslide-prone locations and are therefore unsuitable for early-warning and risk-reduction contexts, where false negatives carry high societal costs. Conversely, metric-driven or density-based losses (e.g., F-measure or mixture density) aggressively optimize training performance but exhibit strong overfitting and poor validation reliability. These findings underline that loss functions must be compatible with both the activation function and the underlying classification objective, rather than optimized in isolation.

Beyond statistical performance, the spatial structure of the SM–ML susceptibility map provides strong physical validation of the optimized configuration. High and very high susceptibility zones are concentrated in steep, deeply dissected terrain characterized by unfavorable combinations of slope angle, fault density, drainage density, rainfall intensity, and anthropogenic disturbance. These zones contain the majority of mapped landslides, whereas very low and low susceptibility classes correspond mainly to gently sloping, stable, and well-vegetated areas with minimal landslide occurrence. Frequency-ratio analysis confirms this stratification, with landslide FR values increasing monotonically from low to high susceptibility classes and non-landslide FR values showing the opposite trend. This clear separation indicates that the probability surface produced by the SM–ML model is not only statistically accurate but also geomorphologically meaningful, which is essential for acceptance in practical hazard management.

Despite these strengths, several limitations remain. Model reliability still depends on the completeness and positional accuracy of the landslide inventory, as well as on the spatial and temporal resolution of conditioning factors, particularly rainfall and land-use data. In this study, network architecture and most hyperparameters were intentionally held constant to isolate the effects of activation and loss functions; future work

could explore additional architectures, regularization strategies, and sampling schemes to further enhance robustness. Incorporating dynamic triggering variables (e.g., rainfall time series, soil moisture indices, near-real-time seismic shaking), testing model transferability to other mountainous regions, and developing ensemble frameworks that combine multiple well-performing activation–loss configurations represent promising directions for future research. Explicit uncertainty quantification would also improve the usefulness of susceptibility maps for risk-informed decision-making.

In short, this study confirms that careful optimization of activation and loss functions, particularly through the SM–ML configuration, offers a simple, computationally efficient, and highly effective pathway to improve deep-learning-based landslide susceptibility assessment. The proposed framework enhances predictive accuracy without increasing model complexity and provides susceptibility maps that are both statistically robust and physically interpretable, supporting more reliable landslide hazard assessment and disaster-risk reduction under changing climatic and land-use conditions.

Acknowledgments

This research was financially supported by the Vietnam Academy of Science and Technology (VAST) under grant code number CBCLCA.16/25-27.

Declaration of Generative AI

During the preparation of this work, the authors used Chat GPT (Open AI) in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- [1] M.J. Froude, D.N. Petley. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18(8), 2161-2181. <https://doi.org/10.5194/nhess-18-2161-2018>
- [2] D. Gómez, E.F. García, E. Aristizábal. (2023). Spatial and temporal landslide distributions using global and open landslide databases. *Natural Hazards*, 117(1), 25-55. <https://doi.org/10.1007/s11069-023-05848-8>
- [3] N.A. Shah, M. Shafique, L.A. Owen, Y. Ai-Mulla, Y. Ullah. (2025). Morphometric analysis of debris flow hazard and risk assessment in the mountain terrains of northern Pakistan using remote sensing and field data. *Earth Science Informatics*, 18(3), 295. <https://doi.org/10.1007/s12145-025-01807-y>
- [4] A. Jaafari. (2024). An Overview of Triggering and Causing Factors of Landslides. in U. Chatterjee, K.C. Lalmalsawmzaupa, B. Biswas, S.C. Pal (Eds.), *Landslides in the Himalayan Region: Risk Assessment and Mitigation Strategy for Sustainable Management*, pp. 25-45. *Springer Nature Singapore*. <https://doi.org/10.1007/978-981-97-4680-4>
- [5] S.T. McColl. (2022). Chapter 2 - Landslide causes and triggers. in: T. Davies, N. Rosser, J.F. Shroder (Eds.), *Landslide Hazards, Risks, and Disasters (Second Edition)*, pp. 13-41. *Elsevier*. <https://doi.org/10.1016/C2018-0-02502-5>
- [6] I. Alcántara-Ayala. (2025). Landslides in a changing world. *Landslides*, 22(9), 2851-2865. <https://doi.org/10.1007/s10346-024-02451-1>
- [7] D.N. Huu, T.V. Cong, P. Bretcan, A.-I. Petrisor. (2024). Assessing the relationship between landslide susceptibility and land cover change using machine learning. *Vietnam Journal of Earth Sciences*, 46(3) 339-359. <https://doi.org/10.15625/2615-9783/20706>
- [8] C.-W. Chen, Y.-S. Tung, F.-Y. Chu, H.-C. Li, Y.-M. Chen. (2025). Assessing landslide risks across varied land-use types in the face of climate change. *Landslides*, 22(9), 2867-2881. <https://doi.org/10.1007/s10346-024-02436-0>
- [9] T.V. Phong, H.-B. Ly, P.T. Trinh, I. Prakash, D.T.

- Hoan. (2020). Landslide susceptibility mapping using Forest by Penalizing Attributes (FPA) algorithm based machine learning approach. *Vietnam Journal of Earth Sciences*, 42(3), 237-246. <https://doi.org/10.15625/0866-7187/42/3/15047>
- [10] H. Bourenane, M.S. Guettouche, Y. Bouhadad, M. Braham. (2016). Landslide hazard mapping in the Constantine city, Northeast Algeria using frequency ratio, weighting factor, logistic regression, weights of evidence, and analytical hierarchy process methods. *Arabian Journal of Geosciences*, 9(2), 154. <https://doi.org/10.1007/s12517-015-2222-8>
- [11] H.M. Do, K.L. Yin, Z.Z. Guo. (2020). A comparative study on the integrative ability of the analytical hierarchy process, weights of evidence and logistic regression methods with the Flow-R model for landslide susceptibility assessment. *Geomatics, Natural Hazards and Risk*, 11(1), 2449-2485. <https://doi.org/10.1080/19475705.2020.1846086>
- [12] A. Ozdemir, T. Altural. (2013). A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *Journal of Asian Earth Sciences*, 64, 180-197. <https://doi.org/10.1016/j.jseaes.2012.12.014>
- [13] H.A.H. Al-Najjar, B. Pradhan, B. Kalantar, M.I. Sameen, M. Santosh, A. Alamri. (2021). Landslide Susceptibility Modeling: An Integrated Novel Method Based on Machine Learning Feature Transformation. *Remote Sensing*, 13(16), 3281. <https://doi.org/10.3390/rs13163281>
- [14] M. Dahim, S. Alqadhi, J. Mallick. (2023). Enhancing landslide management with hyper-tuned machine learning and deep learning models: Predicting susceptibility and analyzing sensitivity and uncertainty. *Frontiers in Ecology and Evolution*, 11, 1108924. <https://doi.org/10.3389/fevo.2023.1108924>
- [15] T.A. Tuan, P.V. Hong, T.T. Tam, N.T.A. Nguyet, N.V. Dung, P.T. Huy, T.V. Phong. (2025). Landslide susceptibility in Phuoc Son, Quang Nam: A deep learning approach. *Vietnam Journal of Earth Sciences*, 47(1), 39-57. <https://doi.org/10.15625/2615-9783/21658>
- [16] T.T.H. Do, D.L. Tran, T.T.H. Nguyen, D.V. Nguyen, V.T. Nguyen. (2025). Application of GIS and remote sensing in detecting coastal reclamation in the coastal areas of Hai Phong - Ha Long from 1990 to 2020. *Vietnam Journal of Marine Science and Technology*, 25(1), 29-39. <https://doi.org/10.15625/1859-3097/20808>
- [17] T.D. Nguyen, Q.M. Ha, K.H. Truong, T.H. Phan, M.H. Cu, V.H. Nguyen. (2024). Applying machine learning for hydraulic flow unit classification and permeability prediction: case study from carbonate reservoir in the Southern part, Song Hong basin. *Vietnam Journal of Marine Science and Technology*, 24(3), 219-234. <https://doi.org/10.15625/1859-3097/18654>
- [18] T.G. Nguyen, T.D. Tran, C.T. Nguyen. (2023). Optimizing the Long Short-Term Memory (LSTM) model by Bayesian method for salinity intrusion forecasting: a study at Dai Ngai station, Soc Trang province, Vietnam. *Vietnam Journal of Marine Science and Technology*, 23(3), 223-232. <https://doi.org/10.15625/1859-3097/18174>
- [19] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C.V. Essen, A.A.S. Awwal, V.K. Asari. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics*, 8(3), 292. <https://doi.org/10.3390/electronics8030292>
- [20] W. Samek, G. Montavon, S. Lapuschkin, C.J. Anders, K.-R. Müller. (2021). Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. *Proceedings of the IEEE*, 109(3), 247-278.

- <https://doi.org/10.1109/JPROC.2021.3060483>
- [21] Y. Li, B. Fu, Y. Yin, X. Hu, W. Wang, W. Wang, X. Li, G. Long. (2024). Review on the artificial intelligence-based methods in landslide detection and susceptibility assessment: Current progress and future directions. *Intelligent Geoengineering*, 1(1), 1-18. <https://doi.org/10.1016/j.ige.2024.10.003>
- [22] J. Hu, Z. Zhang, X. Zhu, X. Zhang, S. Yang, C. Huang, W. Wang, X. Li, L. Hou, L. Zhao. (2025). Geological hazard susceptibility assessment and forecasting analysis based on InSAR and C-L-A model. *International Journal of Applied Earth Observation and Geoinformation*, 143, 104840. <https://doi.org/10.1016/j.jag.2025.104840>
- [23] Q. Zhang, T. Wang. (2024). Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities. *Remote Sensing*, 16(8), 1344. <https://doi.org/10.3390/rs16081344>
- [24] N.D. Anh, T.Q. Cuong, N.C. Quan, N.T. Thanh, T.T. Hieu, B.P. Thao, P.T. Trinh, T.V. Phong, V.C. Dat, I. Prakash, B.T. Pham. (2024). Spatial Prediction of Landslide Susceptibility using Various Machine Learning Based Binary Classification Methods. *Journal of the Geological Society of India*, 100(10), 1477-1492. <https://doi.org/10.17491/jgsi/2024/174002>
- [25] T.X. Bien, P.T. Truyen, T.V. Phong, D.D. Nguyen, M. Amiri, R. Costache, D.M. Duc, H.V. Le, H.B.T. Nguyen, I. Prakash, B.T. Pham. (2022). Landslide susceptibility mapping at sin Ho, Lai Chau province, Vietnam using ensemble models based on fuzzy unordered rules induction algorithm. *Geocarto International*, 37(27), 17777-17798. <https://doi.org/10.1080/10106049.2022.2136253>
- [26] S. Dargan, M. Kumar, M.R. Ayyagari, G. Kumar. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27, 1071-1092. <https://doi.org/10.1007/s11831-019-09344-w>
- [27] B. Aslam, A. Zafar, U. Khalil. (2021). Development of integrated deep learning and machine learning algorithm for the assessment of landslide hazard potential. *Soft Computing*, 25, 13493-13512. <https://doi.org/10.1007/s00500-021-06105-5>
- [28] S. Kiliçarslan, M. Celik. (2021). RSigELU: A nonlinear activation function for deep neural networks. *Expert Systems with Applications*, 174, 114805. <https://doi.org/10.1016/j.eswa.2021.114805>
- [29] M. Goyal, R. Goyal, P.V. Reddy, B. Lall. (2020). Activation Functions. in: W. Pedrycz, S.-M. Chen (Eds.), *Deep Learning: Algorithms and Applications*, pp. 1-30. *Springer, Cham*. https://doi.org/10.1007/978-3-030-31760-7_1
- [30] I. Kouretas, V. Paliouras. (2019). Simplified Hardware Implementation of the Softmax Activation Function. *2019 8th International Conference on Modern Circuits and Systems Technologies (MOCASST)*, pp. 1-4. <https://doi.org/10.1109/MOCASST.2019.8741677>
- [31] S. Scardapane, M. Scarpiniti, D. Comminiello, A. Uncini. (2019). Learning Activation Functions from Data Using Cubic Spline Interpolation. in: A. Esposito, M. Faundez-Zanuy, F.C. Morabito, E. Pasero (Eds.), *Neural Advances in Processing Nonlinear Dynamic Signals*, pp. 73-83. *Springer, Cham*. https://doi.org/10.1007/978-3-319-95098-3_7
- [32] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, E.A. Chávez-Urbiola. (2025). A comprehensive survey of loss functions and metrics in deep learning. *Artificial Intelligence Review*, 58, 195. <https://doi.org/10.1007/s10462-025-11198-7>
- [33] O. Elharrouss, Y. Mahmood, Y. Bechqito, M.A. Serhani, E. Badidi, J. Riffi, H. Tairi. (2025).

- Task-based Loss Functions in Computer Vision: A Comprehensive Review. *arXiv*, *arXiv:2504.04242*.
<https://doi.org/10.48550/arXiv.2504.04242>
- [34] G. Aslan, M. Foumelis, D. Raucoules, M.D. Michele, S. Bernardie, Z. Cakir. (2020). Landslide Mapping and Monitoring Using Persistent Scatterer Interferometry (PSI) Technique in the French Alps. *Remote Sensing*, 12(8), 1305.
<https://doi.org/10.3390/rs12081305>
- [35] N. Casagli, E. Intrieri, V. Tofani, G. Gigli, F. Raspini. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. *Nature Reviews Earth & Environment*, 4, 51-64.
<https://doi.org/10.1038/s43017-022-00373-x>
- [36] I. Goodfellow, Y. Bengio, A. Courville. (2016). Deep Learning. ISBN: 9780262035613. *MIT Press*.
- [37] B. Asadi, H. Jiang. (2020). On Approximation Capabilities of ReLU Activation and Softmax Output Layer in Neural Networks. *arXiv*, *arXiv:2002.04060*.
<https://doi.org/10.48550/arXiv.2002.04060>