

Leveraging geospatial foundation model embeddings for local-scale landslide susceptibility mapping via positive-unlabeled learning

Mingxuan Li

Sun Yat-Sen University

Wenkai Li

liwenk3@mail.sysu.edu.cn

Sun Yat-Sen University <https://orcid.org/0000-0001-6548-7882>

Hongliang Liu

Sun Yat-Sen University

Changshan Liang

Sun Yat-Sen University

Research Article

Keywords: Geospatial foundation model, Landslide susceptibility mapping, AlphaEarth, Positive-unlabeled learning, Hong Kong

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10786241/v1>

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

[Read Full License](#)

1 **Leveraging geospatial foundation model embeddings for local-scale**
2 **landslide susceptibility mapping via positive-unlabeled learning**

3
4 **Mingxuan Li¹, Wenkai Li^{1,*}, Hongliang Liu¹, Changshan Liang¹**

5
6 ¹ School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, China

7
8 * **Corresponding author:** Wenkai Li (email: liwenk3@mail.sysu.edu.cn; ORCID: 0000-0001-
9 6548-7882)

10
11 **Acknowledgments**

12 The authors gratefully acknowledge the Civil Engineering and Development Department of
13 Hong Kong for providing the open-access historical landslide dataset Enhanced Natural Terrain
14 Landslide Inventory (ENTLI).

Abstract

Geospatial foundation models (GeoFMs) provide transferable embeddings that may reduce manual feature engineering in landslide susceptibility mapping (LSM). However, their application to local-scale LSM remains constrained by unreliable negative samples and redundancy in high-dimensional representations. This study proposes a deployment-oriented workflow that integrates AlphaEarth embeddings with positive-unlabeled (PU) learning, thereby aligning transferable geospatial representations with the intrinsic positive-unlabeled structure of landslide inventories for practical local-scale LSM, using the Hong Kong landslide inventory as a testbed. Methodologically, we first diagnosed and mitigated multicollinearity within embeddings and traditional physical features through targeted optimization, including dimensionality reduction and regularization. We then trained PU classification models using the Positive and Background Learning with Constraints algorithm under three feature configurations: Embedding-only, Traditional-only, and Fused. Furthermore, we evaluated model interpretability and decision stability using SHapley Additive exPlanations and rank-correlation metrics. The results show that the Embedding-only configuration provides a competitive, low-preprocessing baseline. When multicollinearity is mitigated, the Fused model achieves the best overall performance and the most spatially selective delineation of high-susceptibility areas, with the majority of landslides captured within only 6.3% of the study area and a landslide density approximately 24 times the background level. The results further suggest that GeoFM embeddings complement traditional physical features by providing fine-grained surficial semantics, helping reduce false positives in complex built environments, and partially substituting for some satellite-derived indices. Overall, this workflow supports rapid

engineering-geological screening and more detailed susceptibility assessment under realistic label constraints.

Keywords: Geospatial foundation model; Landslide susceptibility mapping; AlphaEarth; Positive-unlabeled learning; Hong Kong

1. Introduction

Landslides are among the most common and destructive natural hazards worldwide, leading to significant loss of life and infrastructure (Haque et al. 2016; Froude and Petley 2018). Landslide susceptibility mapping (LSM) aims to predict future landslides by analyzing the relationship between historical landslide inventories and environmental factors, and thus plays a crucial role in disaster management (Guzzetti et al. 1999). For engineering-geological practice, local-scale LSM supports rapid hazard screening, pre-event preparedness, and the prioritization of hillside management, land-use planning, and field investigation (Reichenbach et al. 2018). However, directly deploying advanced artificial intelligence, such as foundation models, for operational local-scale LSM remains challenging. The primary bottlenecks lie in effective feature representation and the reliability of training labels.

From the perspective of feature representation, conventional LSM typically relies on physically interpretable conditioning factors, such as Elevation, Slope, Normalized Difference Vegetation Index, Topographic Wetness Index, and Stream Power Index (Meena et al. 2022). While these factors are widely used, they often struggle to capture subtle surface textures and complex non-linear interactions in heterogeneous terrains, which can limit predictive accuracy

(Merghadi et al. 2020). While deep learning algorithms such as convolutional neural networks provide stronger feature extraction capabilities to overcome this limitation (Liu et al. 2022; Chen et al. 2025b), they inadvertently trigger the second bottleneck: the demand for large labeled datasets to train these models from scratch, which is often unfeasible in local operational LSM (Jiang et al. 2025).

The geospatial foundation model (GeoFM) paradigm offers a promising alternative. By leveraging self-supervised pretraining on massive Earth observation data, GeoFM can extract transferable representations that are readily adapted to downstream tasks with sparse labels (Mai et al. 2024). For example, the Prithvi model, trained from Landsat and Sentinel-2 data (Szwarcman et al. 2026), significantly outperforms conventional models in LSM (Li et al. 2025). General-purpose GeoFMs such as SkySense (Guo et al. 2024), TerraMind (Jakubik et al. 2025), and StefaLand (Kraabel et al. 2025) have demonstrated strong performance across a broad range of Earth observation tasks. More task-specific studies have begun to explore GeoFM representations for hazard applications, including flood hazard mapping (Kim and Oikonomou 2025) and landslide hazard mapping (Li et al. 2025). By leveraging transformer architectures and large-scale pretraining, GeoFMs have shown strong potential in terms of predictive accuracy, data efficiency, and cross-sensor generalization across Earth observation tasks (Reed et al. 2023; Hong et al. 2024). However, deploying these models often requires complex fine-tuning or is limited to specific sensor inputs and scope. In contrast, AlphaEarth offers a more deployment-friendly approach. Backed by comprehensive global coverage, it directly generates compact embeddings from multi-source imagery, implicitly encoding cumulative spatio-temporal surface information with minimal preprocessing (Brown et al.

2025). This “ready-to-use” property makes AlphaEarth particularly attractive for operational LSM, as it reduces manual feature engineering while providing continuous, fine-grained semantic descriptors beyond traditional physical factors.

Nevertheless, a key question remains: how can GeoFM be reliably deployed in local-scale LSM? Reliable deployment at the local scale raises at least two practical questions. First, it is unclear whether GeoFM embeddings alone can match the explanatory power of traditional physical factors, and whether they should be used as a standalone representation or fused with traditional physical factors. Second, GeoFM embeddings are typically high-dimensional and may exhibit substantial redundancy and multicollinearity due to spatial autocorrelation when applied to regional datasets, which can compromise model robustness and interpretability if used without targeted optimization (Dormann et al. 2013).

In parallel, the label structure of landslide inventories introduces another major barrier to operational LSM. In practice, reliable negative samples (non-landslides) are seldom available, as the absence of a record does not guarantee hillside stability; it may simply reflect incomplete surveys or a lack of triggering events (Wang et al. 2022; Dou et al. 2023). Consequently, conventional Positive–Negative (PN) sampling often selects negative samples based on subjective rules or simplified assumptions, introducing biased or noisy negatives that degrade model generalization (Chen et al. 2025a). Positive-Unlabeled (PU) learning instead trains on confirmed landslides and unlabeled background samples, allowing a larger pool of observations to contribute to model training without requiring additional negative annotation (Bekker and Davis 2020; Fu et al. 2024). This property is particularly useful in the present study because each observation is represented by a high-dimensional GeoFM embedding. PU-

based LSM has been explored in various forms, including PU-bagging (Wu et al. 2021), two-step strategies (Yao et al. 2022) and other PU-enhanced classifiers (Zhang et al. 2024) for handling unlabeled data. However, most PU-based susceptibility studies still rely on traditional physical factors, and the integration of PU learning with GeoFM embeddings remains underexplored.

Motivated by these gaps, this study investigates a practical pathway for deploying GeoFM in LSM by combining (1) a PU learning paradigm that matches the intrinsic label structure of landslide inventories, and (2) a systematic evaluation and optimization mechanism for high-dimensional embedding features. Using Hong Kong as a case study, we adopt the Positive and Background Learning with Constraints (PBLC) algorithm as a PU learning backbone and examine three feature configurations: Embedding-only, Traditional-only, and Fused feature sets. Based on performance comparison, interpretability analysis, and stability diagnostics, this work provides empirical evidence and practical guidance for deploying GeoFM embeddings in local-scale LSM.

Specifically, the contributions of this research are threefold:

(1) **A PU-learning deployment strategy for engineering-geological susceptibility assessment under unreliable negative labels.** We adopt a PBLC-based modeling workflow that learns from positive landslide inventories and unlabeled data, mitigating bias introduced by negative sampling and supporting rapid deployment without manually constructing reliable negative samples.

(2) **A systematic benchmark of GeoFM embeddings in local-scale LSM.** We compare Embedding-only, Traditional-only, and Fused feature sets to quantify whether GeoFM

embeddings are sufficient as a baseline and how much additional gain can be achieved by fusing embeddings with traditional physical factors.

(3) **Practical guidelines for using high-dimensional embeddings robustly.** We explicitly diagnose redundancy and multicollinearity in GeoFM embeddings and evaluate optimization strategies such as dimensionality reduction and regularization, to improve model robustness and maintain interpretable decision logic when embeddings are applied alone or fused with traditional physical factors.

2. Methodology

To deploy GeoFM embeddings for local-scale LSM under a PU setting, we developed a systematic workflow that combines AlphaEarth-derived embeddings with PBLC-based modeling, using Hong Kong as a case study. The methodology was structured into three core phases: (1) diagnosing and mitigating multicollinearity in the high-dimensional embedding space for robust representation; (2) modeling and benchmarking landslide susceptibility across Embedding-only, Traditional-only, and Fused feature configurations; and (3) conducting interpretability and stability analyses via SHapley Additive exPlanations (SHAP) and rank-correlation to evaluate the model's decision logic. Ultimately, these steps establish a concrete empirical foundation for using GeoFM in local-scale hazard assessment.

2.1. Study area and landslide inventory

Hong Kong (HK) is located on the southeastern side of the Pearl River Delta and comprises Hong Kong Island, the Kowloon Peninsula, and the New Territories (including Lantau Island

and over 260 outlying islands). The region is characterized by a complex, rugged mountainous terrain. The scarcity of flat terrain has necessitated high-density urban development directly adjacent to these steep foothills (Wong 2009). Governed by a subtropical monsoon climate, HK experiences hot, humid summers with frequent typhoons and extreme rainfall (Ko and Lo 2016). The convergence of these complex geological conditions and seasonal intense precipitation renders HK one of the most landslide-prone urban areas globally.

Historical landslide data were collected from the Enhanced Natural Terrain Landslide Inventory (ENTLI), published by the Civil Engineering and Development Department of Hong Kong. This dataset, constructed through high-resolution aerial photograph interpretation and field investigations, details the spatial distribution and geometry of natural terrain landslides (Gao et al. 2021). These recorded landslide events were directly selected as positive samples. To ensure spatio-temporal consistency between landslide events and predictor variables, the following screening criteria were applied: (1) Temporal Filtering: Only records from 2017 onwards were retained to align with the acquisition period of AlphaEarth; and (2) Spatial Filtering: Samples falling outside the study area or lacking complete environmental data coverage were excluded. This process resulted in a final dataset of 1059 historical landslide events (Fig. 1).

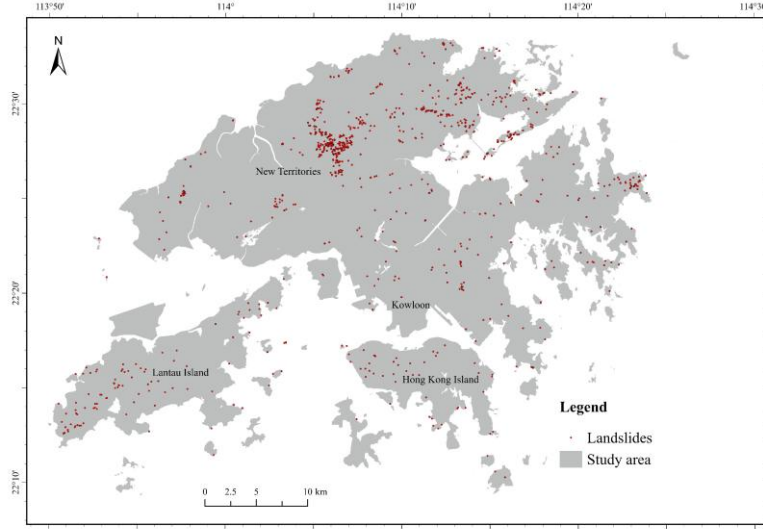


Fig. 1 Boundary of the study area and spatial distribution of the landslide inventory in HK (map lines delineate study areas and do not necessarily depict accepted national boundaries)

2.2. Predictor variables and preprocessing

To comprehensively characterize the environmental conditions conducive to landslides, a total of 81 predictor variables were derived from multi-source data. The detailed data sources and extracted variables are summarized in Table 1. These variables were categorized into six groups based on their attributes: (1) GeoFM Embeddings; (2) Topographical Factors; (3) Hydrological Factors; (4) Geological Factors; (5) Surface Activity Factors; and (6) Satellite-derived Factors. The spatial distribution of representative embedding features and traditional physical factors is visualized in Fig. 2. All spatial data were resampled to a 10 m resolution. Categorical variables, including Aspect, Lithology, Land Use, and Land Cover, were numerically encoded using the frequency ratio method to meet the input requirements of the model:

$$T(x_i) = \frac{\Pr(x_i|y=1)}{\Pr(x_i)} \quad (1)$$

where $T(x_i)$ represents the transformed value for the i -th category of variable x ; $\Pr(x_i | y = 1)$

denotes the frequency of the i -th category within landslide locations; and $\Pr(x_i)$ represents the background frequency of the i -th category across the entire study area (Li et al. 2022).

Table 1. Overview of data sources and extracted predictor variables.

Dataset	Features	Provider	Type	Resolution
Google Satellite Embedding	Embeddings (64 bands): A00 to A63	Google Earth Engine	Raster	10 m
Digital Terrain Model	Elevation, Slope, Aspect, Plan Curvature, Profile Curvature, TWI, SPI	HK Lands Department	Raster	5 m
Road Centreline	Distance to Roads	HK Lands Department	Vector	-
Water Framework Spatial Data	Distance to Rivers	HK Water Supplies Department	Vector	-
GEO Raingauge Data	Average Daily Precipitation	HK Civil Engineering and Development Department	Vector	-
Geological Maps	Lithology, Distance to Faults	HK Civil Engineering and Development Department	Vector	-
Population Census	Population Density	HK Census and Statistics Department	Vector	-
Land Utilization	Land Use	HK Planning Department	Raster	10 m
ESRI 10m Annual Land Cover	Land Cover	Google Earth Engine	Raster	10 m
Sentinel-2	NDVI, NDBI	Google Earth Engine	Raster	10 m

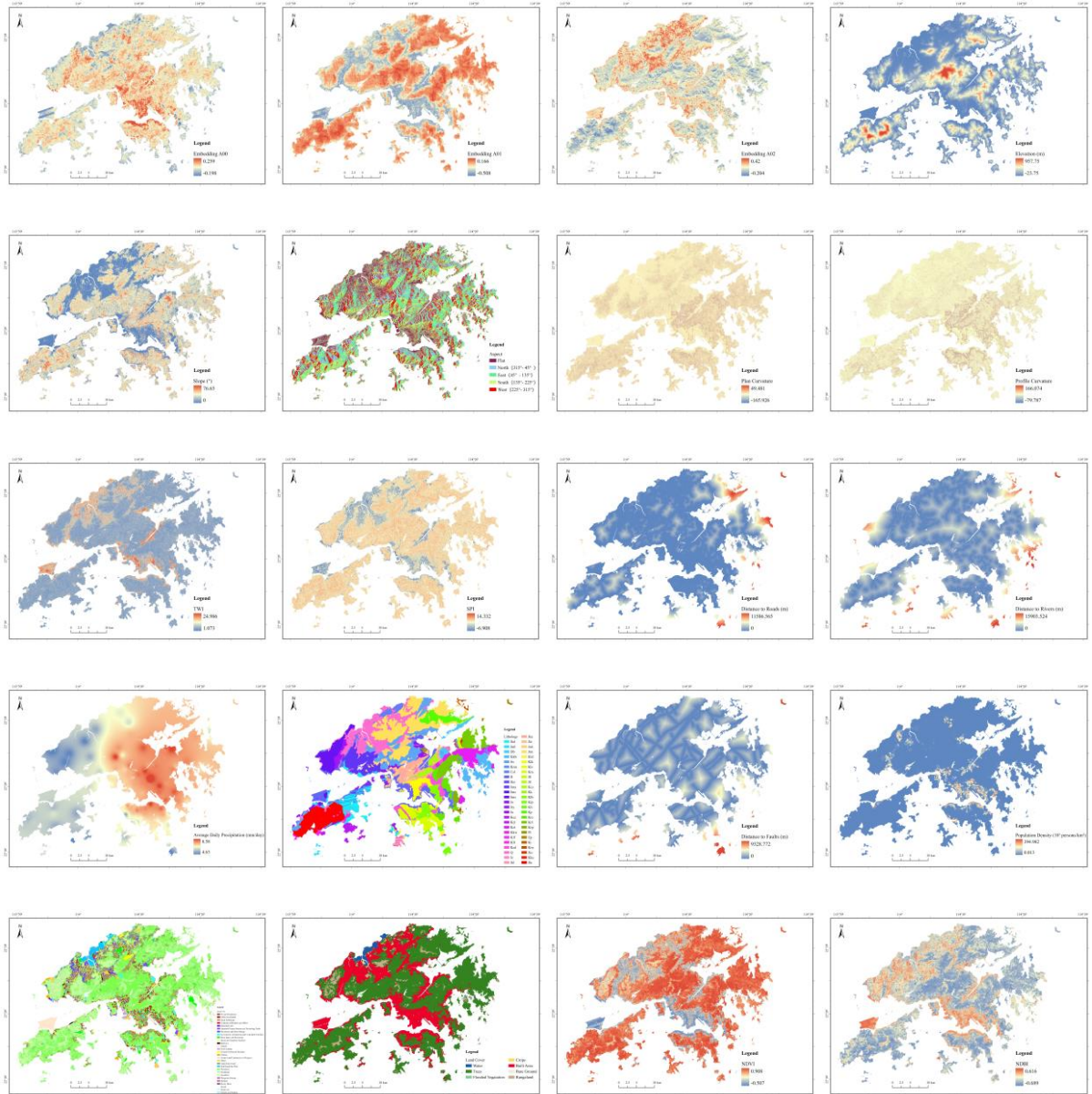


Fig. 2 Spatial visualization of representative embedding features (e.g., A00, A01, and A02) and traditional physical factors

GeoFM Embeddings. High-dimensional environmental semantic information was extracted using the AlphaEarth Model. To aggregate multi-temporal information, we calculated the time-series mean of the AlphaEarth embeddings from 2017 to 2024, resulting in 64 embedding features (**A00–A63**). Unlike traditional physical factors, these embeddings are designed to capture fine-grained surface textures and latent semantic patterns that are difficult

to quantify through conventional means (Brown et al. 2025).

Topographical Factors. Key geomorphological parameters were extracted from a 5 m resolution Digital Terrain Model (DTM), including **Elevation, Slope, Aspect, Plan Curvature,** and **Profile Curvature**. Aspect was initially reclassified into five categories (North, South, East, West, and Flat) and subsequently transformed into continuous values using Eq. 1.

Hydrological Factors. Given that hydrological conditions govern pore water pressure within hillsides, the **Topographic Wetness Index (TWI)** and **Stream Power Index (SPI)** were calculated to characterize topographic control on saturation and erosion potential:

$$TWI = \ln\left(\frac{SCA}{\tan \beta}\right) \quad (2)$$

$$SPI = SCA \times \tan \beta \quad (3)$$

where *SCA* denotes the Specific Catchment Area and β represents the slope (Li et al. 2022). Additionally, the **Distance to Rivers** was calculated, and an **Average Daily Precipitation** distribution map was generated via inverse distance weighting interpolation based on rain gauge observations from 2017–2024.

Geological Factors. Geological structures provide the material basis for hillsides. **Lithology** was extracted from geological maps and encoded via Eq. 1, while **Distance to Faults** was calculated using Euclidean distance to quantify the impact of tectonic activity on rock mass integrity.

Surface Activity Factors. This category reflects the influence of anthropogenic disturbances. For **Land Use** and **Land Cover**, the mode category for the periods 2018–2023 and 2017–2024, respectively, was selected and transformed via Eq. 1. **Distance to Roads** was included to reflect disturbances to hillside stability caused by road construction. **Population**

Density, derived from 2021 census polygon data, was converted to a raster and spatially smoothed to ensure continuous transitions across census district boundaries, providing a more realistic representation of human activity.

Satellite-derived Factors. Vegetation and built-up indices were calculated using Sentinel-2 multispectral imagery. The **Normalized Difference Vegetation Index (NDVI)** and **Normalized Difference Built-up Index (NDBI)** were derived from the time-series mean for 2017–2024. To maintain spatial consistency, the 20 m resolution short-wave infrared band (B11) used for NDBI was resampled to 10 m using nearest-neighbor interpolation.

To mitigate redundancy within the feature space, a preliminary collinearity diagnostic protocol was established using the Variance Inflation Factor (VIF) and Pearson correlation coefficient, based on 5295 random samples (five times the number of positive samples). Following this screening protocol, three distinct feature sets were defined: **Embedding-only** (GeoFM embeddings only), **Traditional-only** (all variables except GeoFM embeddings), and **Fused** (all variables combined). After the variable selection process, the remaining features underwent Z-score normalization. To address severe internal multicollinearity, a Principal Component Analysis (PCA) layer retaining 95% of the cumulative variance was implemented for the Embedding-only and Fused sets. Conversely, PCA was omitted for the Traditional-only set to preserve the explicit physical interpretability of these low-dimensional, pre-screened factors. PCA was implemented as an integrated preprocessing layer within the neural network architecture.

2.3. Landslide susceptibility modeling and assessment

To address the inherent PU characteristics of landslide inventory data, this study adopts the PBLC algorithm to construct the predictive model. Unlike traditional supervised learning, which requires labeled negative samples, PBLC uses labeled positive samples together with unlabeled background samples as model input. PBLC estimates the posterior probability of landslide occurrence, $\Pr(y = 1 | x)$, where x denotes the predictor vector and $y = 1$ indicates landslide occurrence. This is achieved by modeling the relationship between the observed label variable s and the true state y , where $s = 1$ denotes a labeled landslide:

$$\Pr(s = 1|x) = \frac{\Pr(y=1|x)}{\Pr(y = 1|x) + \frac{1-c}{c}} \quad (4)$$

where $c = \Pr(s = 1 | y = 1)$ denotes the label frequency, representing the probability that a landslide is recorded (labeled). By introducing a maximum probability constraint into the loss function, specifically assuming the maximum probability of landslide occurrence in the study area is $P_{max} \approx 1$, the PBLC algorithm iteratively solves for the parameter c during the training process (Li et al. 2021).

A fully connected Artificial Neural Network (ANN) serves as the base classifier to fit these probabilities. The network architecture comprises three hidden layers with 32, 16, and 8 neurons, respectively. To enhance generalization, each hidden layer is followed by a Batch Normalization layer and a Dropout layer (rate = 0.3).

To simultaneously address positive-negative imbalance, label noise, and overfitting, a composite loss function, L_{total} , was constructed. First, Focal Loss L_{focal} was employed. By setting the focusing parameter $\gamma = 2.0$, the model reduces the weight of easily classified background samples while intensifying focus on hard-to-classify positive samples:

$$L_{focal} = -\frac{1}{N} \sum_{i=1}^N [s_i(1 - \hat{s}_i)^\gamma \log(\hat{s}_i) + (1 - s_i)\hat{s}_i^\gamma \log(1 - \hat{s}_i)] \quad (5)$$

where N is the total number of samples in the batch, $s_i \in \{0, 1\}$ represents the observed label of the i -th sample, and $\hat{s}_i$ denotes the predicted probability of the observed label (Lin et al. 2017).

Second, based on the prior assumption that $P_{max} \approx 1$, a PBLC constraint term ($L_{constraint}$) is incorporated. Using a regularization coefficient $\lambda = 0.1$, the model penalizes the deviation of the maximum predicted true posterior probability from this prior expectation:

$$L_{constraint} = \lambda [\max(\Pr(y = 1|x)) - P_{max}]^2 \quad (6)$$

Combined with an L2 regularization term (L_{l2}) to restrict model complexity, the total optimization objective is defined as:

$$L_{total} = L_{focal} + L_{constraint} + L_{l2} \quad (7)$$

In the experimental setup, a positive-to-background sample ratio of 1:5 was adopted empirically, with background points randomly sampled from the study area. Samples were randomly partitioned into training, validation, and testing sets using a 7:2:1 split. To mitigate uncertainties from data sampling and model initialization, a repeated evaluation strategy was implemented. Specifically, 10 independent datasets were generated. For each dataset, the entire pool of positive samples was fixed while a new batch of background samples was randomly drawn. Both positive and background samples were then randomly partitioned into training, validation, and testing sets. Subsequently, the model was independently trained 10 times on each dataset, yielding a total of 100 experimental runs per feature set combination. Training utilized the Adam optimizer with a batch size of 512 and an initial learning rate of 0.001, which was dynamically adjusted via a cosine decay schedule with a 20-epoch restart cycle. An early stopping mechanism was applied, terminating training if the validation loss failed to decrease

for 20 consecutive epochs after an initial 80-epoch warmup.

Model performance was evaluated using specialized metrics for PU learning environments, including the Area Under the ROC Curve (AUC) and the F_{pb} score, which serves as a proxy for the F_1 -score in the absence of reliable negative labels:

$$F_{pb} = \frac{2 \times TP}{TP + FN + FP} \quad (8)$$

where TP represents the number of correctly predicted positive samples, FN is the number of missed positive samples, and FP denotes background samples predicted as landslides. The F_{pb} score reflects the model's capacity to balance the effective identification of historical landslides against the suppression of false alarms in background areas, thereby indicating its capacity for more spatially precise prediction (Li and Guo 2021).

2.4. Feature importance and its stability

To improve the interpretability of the deep learning model and quantify the marginal contribution of each predictor variable, this study employs the SHAP method. Rooted in cooperative game theory, SHAP decomposes the model's predictive output into additive contributions attributable to each input feature. For a given test sample, the SHAP value assigned to a feature represents the magnitude and direction of the feature's influence relative to the model's baseline prediction. Unlike traditional sensitivity analysis, SHAP effectively captures non-linear relationships and complex interactions among features (Lundberg and Lee 2017).

This interpretability is particularly crucial for the 64-dimensional GeoFM embeddings. Due to their abstract mathematical nature, these embeddings lack intuitive physical definitions.

By analyzing global SHAP importance rankings, we aim to determine the actual contribution of these high-dimensional features to the model’s decision-making process. This analysis is essential to ascertain whether GeoFM embeddings provide unique environmental semantics beyond the reach of traditional physical factors, rather than merely introducing computational redundancy and noise.

Beyond the magnitude of feature importance, this study also evaluates the reliability of these rankings. While high-dimensional embeddings enrich the feature space, they may also increase model sensitivity to minor perturbations in training data, potentially exacerbating overfitting risks. To assess this, Spearman’s Rank Correlation Coefficient was employed to quantify the consistency of feature importance rankings across experiments (Khair and Dhanalakshmi 2022). By calculating pairwise correlations across the 100 independent experimental runs (as described in Section 2.3), we objectively assessed whether the model maintains robust decision logic under data perturbations. A consistently high coefficient value indicates that the model maintains relatively stable feature-importance rankings across repeated experiments, thereby supporting the robustness of its decision logic (Stow and Stewart 2025).

3. Results

3.1. Multicollinearity diagnostics and feature set optimization

Feature redundancy and multicollinearity can strongly affect model predictive quality. Fig. 3a presents the hierarchical clustering heatmap based on Pearson correlation coefficients. The results reveal distinct block structures, particularly within the 64 embedding features, which form dense clusters of internal correlation. This indicates substantial redundancy within the

GeoFM embedding space.

Network topology (Fig. 3b) and VIF statistics (Fig. 3c) provide a comprehensive diagnostic for feature selection. First, regarding satellite-derived indices, the network graph (Fig. 3b) reveals that NDVI and NDBI form strong linkages ($|R| > 0.75$) with specific embedding features. This pattern indicates that NDVI and NDBI are strongly associated with specific GeoFM embedding dimensions.

In contrast, topographical factors exhibit a unique pattern. Although Elevation displays high VIF values (Fig. 3c), it remains structurally isolated from the embeddings in the network graph (Fig. 3b), with no strong linear connections to GeoFM embeddings. This decoupling indicates that Elevation captures information not strongly represented in the GeoFM embedding. Therefore, despite its high VIF, Elevation was retained because it showed limited direct linear overlap with GeoFM embeddings in the correlation network.

Finally, regarding the embedding features themselves, Fig. 3c confirms severe internal multicollinearity, with the median VIF significantly exceeding the threshold of 10 and individual dimensions (e.g., A08) reaching as high as 84.31. Such strong internal redundancy suggests that directly inputting raw embeddings could complicate model optimization and undermine interpretability.

Guided by these quantitative diagnostics, the final feature sets were screened and restructured. Specifically, NDVI and NDBI were explicitly excluded due to their demonstrated redundancy with the GeoFM embeddings. For the retained high-dimensional embeddings, PCA was applied to retain orthogonal components explaining 95% of the cumulative variance, thereby providing more stable inputs for the subsequent ANN model.

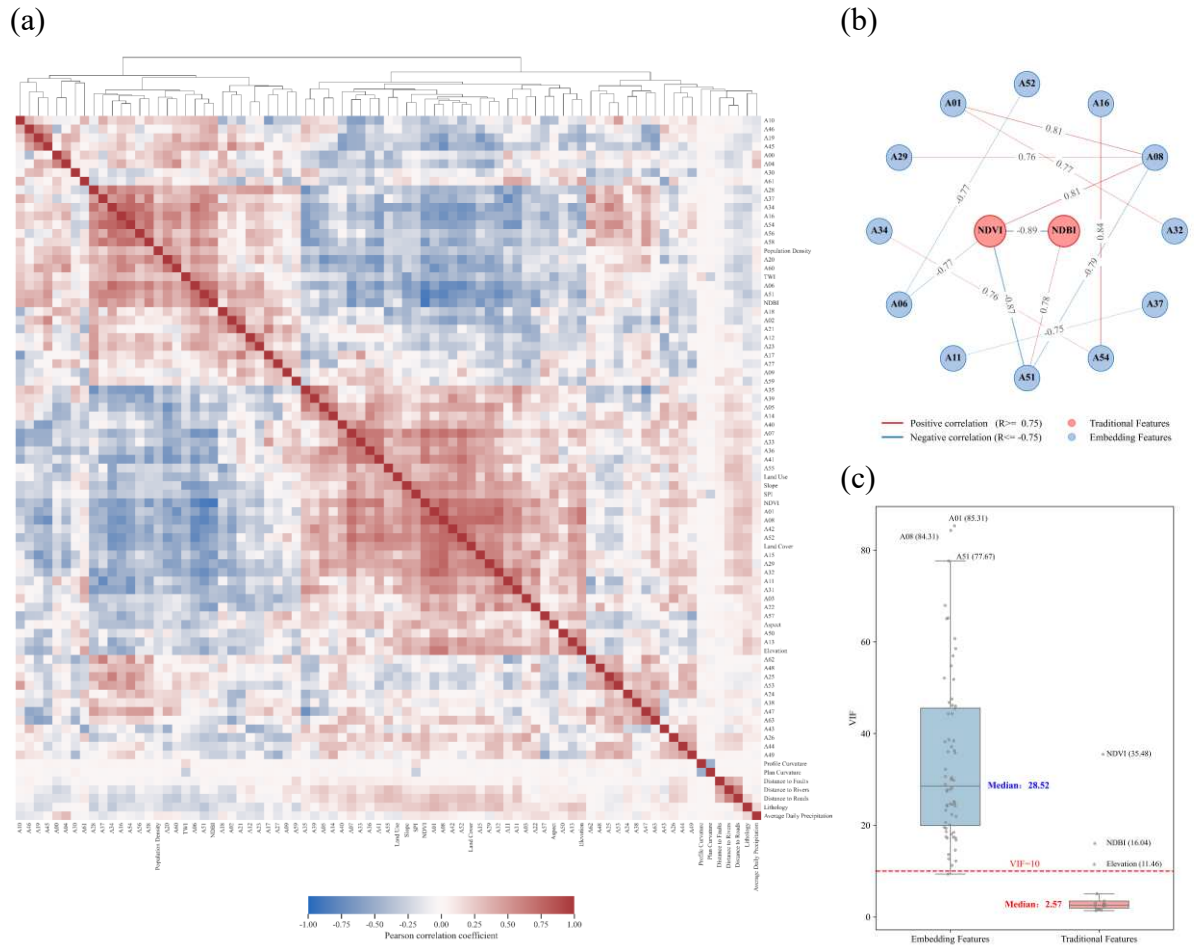


Fig. 3 Multicollinearity analysis of predictor variables: (a) Pearson correlation matrix heatmap with hierarchical clustering; (b) Network graph highlighting strong linear relationships ($|R| > 0.75$); (c) Distribution of VIF values grouped by embedding and traditional physical features (calculated jointly across the Fused feature set)

3.2. Comparative assessment and spatial mapping of landslide susceptibility

The quantitative assessment (Fig. 4) demonstrates that different feature sets have a substantial impact on model performance. The Embedding-only model yields the lowest median AUC and F_{pb} values and shows the widest interquartile ranges, indicating comparatively lower and less stable performance. In contrast, the Traditional-only model significantly outperforms the Embedding-only model across all metrics, leveraging the strong explanatory power of explicit

physical factors.

The Fused model demonstrates the dual advantages of superior precision and robust stability. As illustrated in the box plots (Fig. 4d), the Fused model not only achieves the highest median AUC and F_{pb} values but also exhibits the narrowest interquartile ranges, indicating the most stable performance among the three configurations. A distinct statistical gap is observed: the lower quartile of the Fused model exceeds the upper quartile of the Embedding-only model. Furthermore, the higher F_{pb} values suggest that the Fused model achieves more spatially focused prediction of historical landslide locations.

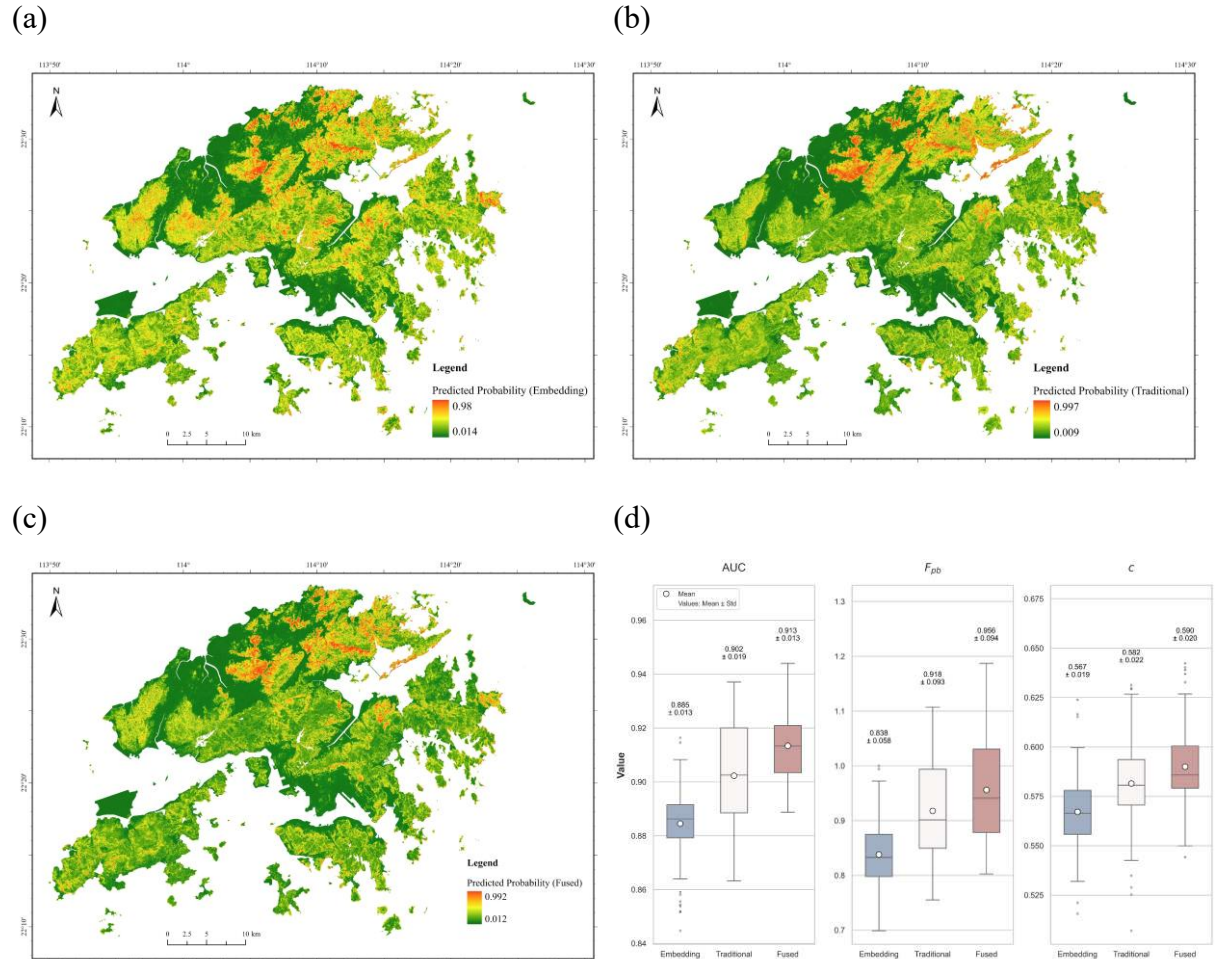


Fig. 4 LSM across different feature sets: (a-c) Susceptibility maps generated by averaging the predicted probabilities from 100 independent experiments for the Embedding-only,

Traditional-only, and Fused models, respectively; (d) Box plots illustrating the distribution of performance metrics (AUC and F_{pb}) aggregated over these 100 runs

Spatially (Fig. 4a-4c), all three models consistently identify the mountainous terrain of the northern New Territories as high-susceptibility zones. However, notable differences emerge in highly urbanized areas such as the Kowloon Peninsula. The Traditional-only model assigns relatively high susceptibility to some elevated roads and other built structures, whereas such patterns are less pronounced in the Embedding-only map. The Fused model retains the major high-susceptibility patterns in mountainous terrain while showing fewer apparent false positives in urban areas.

Statistical analysis of susceptibility classes (Table 2) further compares the spatial concentration patterns of the three models. The Fused model delineates the most spatially concentrated high-susceptibility zones. Its high-susceptibility zones (probability > 0.6) occupy only 6.3% of the study area, compared with 9.1% for the Embedding-only model and 7.3% for the Traditional-only model. At the same time, only 6.9% of positive samples fall into low-susceptibility zones (probability < 0.4), compared with 8.0% and 13.4% for the Embedding-only and Traditional-only models, respectively. In the highest susceptibility interval (0.8 - 1.0), the Fused model reaches a frequency ratio of 24.736, corresponding to a landslide density about 24 times higher than the background average. These results indicate that the Fused model concentrates historical landslides more effectively into a smaller high-susceptibility area than the other two models.

Table 2. Statistics of landslide susceptibility classes, landslide density, and frequency ratios across the three feature sets, derived from the ensemble average of 100 independent experiments.

Feature Set	Probability Interval	Grid Count	Landslide Count	Grid Ratio	Landslide Ratio	Frequency Ratio
Embedding-only	0.0 - 0.2	4087513	1	0.386	0.001	0.002
	0.2 - 0.4	3815076	84	0.360	0.079	0.220
	0.4 - 0.6	1717881	239	0.162	0.226	1.390
	0.6 - 0.8	769675	406	0.073	0.383	5.272
	0.8 - 1.0	193014	329	0.018	0.311	17.034
Traditional-only	0.0 - 0.2	4895932	3	0.463	0.003	0.006
	0.2 - 0.4	3582007	139	0.338	0.131	0.388
	0.4 - 0.6	1329428	233	0.126	0.220	1.752
	0.6 - 0.8	533521	270	0.050	0.255	5.057
	0.8 - 1.0	242271	414	0.023	0.391	17.077
Fused	0.0 - 0.2	6206567	3	0.586	0.003	0.005
	0.2 - 0.4	2620913	70	0.248	0.066	0.267
	0.4 - 0.6	1092241	242	0.103	0.229	2.214
	0.6 - 0.8	507895	359	0.048	0.339	7.064
	0.8 - 1.0	155543	385	0.015	0.364	24.736

3.3. Feature interpretability and model stability analysis

The SHAP analysis reveals distinct feature-contribution patterns across the different feature combinations (Fig. 5). The global importance rankings indicate that explicit physical factors remain dominant predictors of landslide susceptibility. In both the Traditional-only and Fused models, Slope, Lithology, and Land Use consistently rank as the top factors.

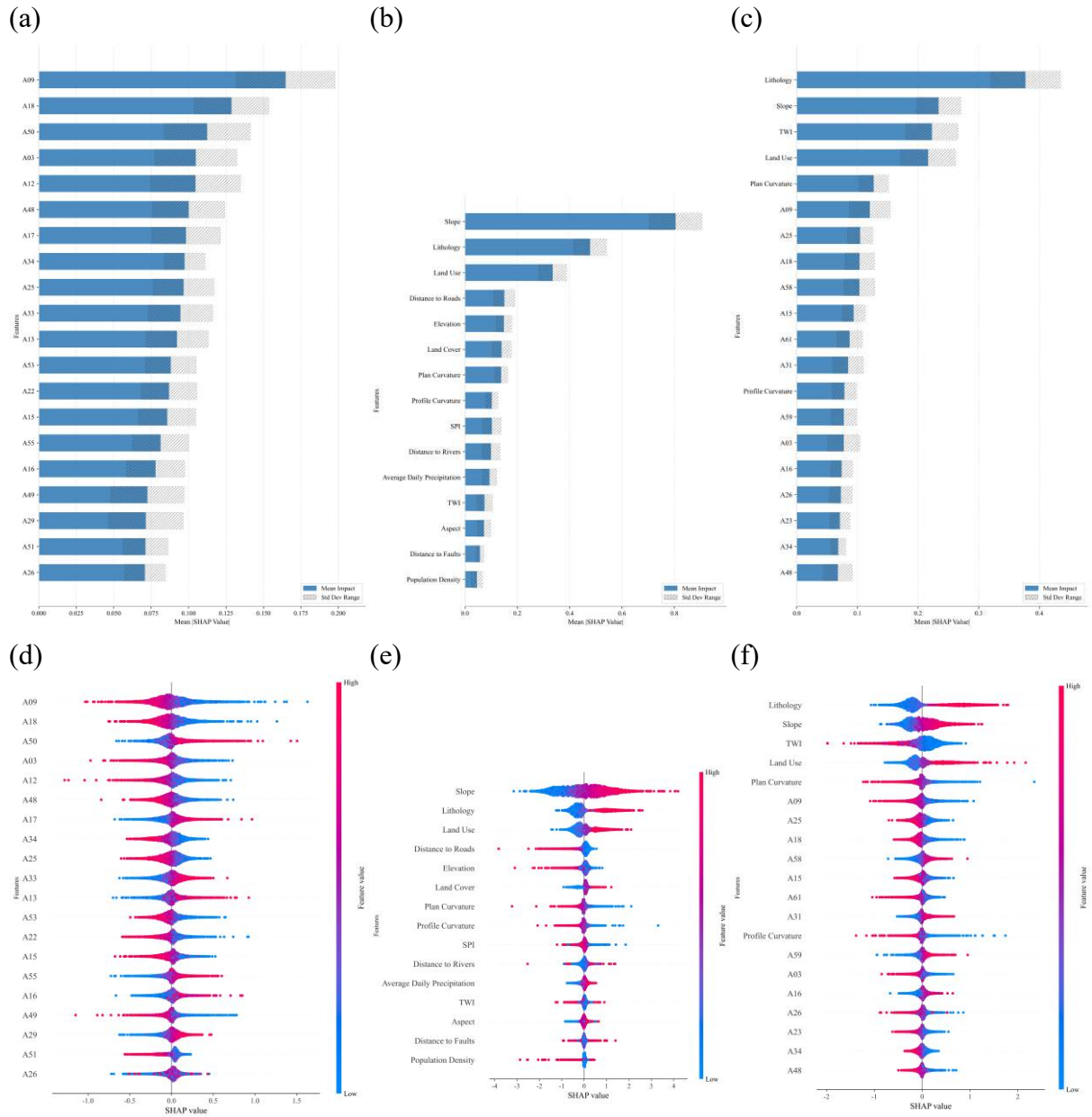


Fig. 5 SHAP-based feature importance analysis: (a-c) Global feature importance rankings (mean |SHAP|), derived by averaging absolute SHAP values across 100 independent experiments for the Embedding-only, Traditional-only, and Fused models; (d-f) SHAP summary plots illustrating the directional impact of feature values on predictive output, constructed from a cumulative dataset of 5,000 observations (50 instances randomly sampled from each of the 100 experimental runs)

A notable shift in feature importance is observed in the Fused model: Lithology becomes

the top-ranked factor, and the importance of TWI also increases. Relative to the Traditional-only model, this pattern suggests that the introduction of GeoFM embeddings changes the weighting of traditional predictors, with greater emphasis placed on geological and hydrological variables.

Regarding the embeddings within the Embedding-only and Fused models, specific top-ranked dimensions (e.g., A09 and A18) are associated with lower predicted susceptibility at higher feature values. Unlike the pronounced dominance of the top three physical factors, the SHAP values within the embedding dimensions are more evenly distributed. Compared with the stronger concentration of importance in a few physical factors, the contribution of embedding dimensions is more evenly distributed, suggesting that no single embedding dimension dominates model prediction.

The ablation results (Table 3) reveal a clear trade-off between predictive performance and decision stability. Specifically, while removing the PCA layer (directly inputting raw features) yields only marginal increases in test-set metrics (evident when comparing experiment No. 2 to No. 1, or No. 5 to No. 4 for Test AUC and F_{pb}), it simultaneously results in a substantial decline in the SHAP Spearman correlation coefficient. Additionally, this removal leads to a widening of both the AUC Gap and F_{pb} Gap in these corresponding scenarios. This pattern suggests that, without dimensionality reduction, high-dimensional multicollinearity is associated with less stable feature-importance rankings and larger train-test performance gaps. Accordingly, the PCA layer appears more valuable for improving robustness and consistency than for maximizing raw predictive accuracy.

Table 3. Ablation results for PCA and regularization across 100 independent runs.

No.	Feature Set	PCA	Reg.	SHAP Spearman	Test AUC	AUC Gap	Test F_{pb}	F_{pb} Gap
1	Fused	✓	✓	0.708 ±0.102	0.913 ±0.013	0.015 ±0.014	0.956 ±0.094	0.052 ±0.114
2	Fused		✓	0.568 ±0.099	0.916 ±0.012	0.028 ±0.014	0.986 ±0.093	0.108 ±0.113
3	Fused			0.282 ±0.101	0.908 ±0.012	0.069 ±0.013	0.870 ±0.142	0.258 ±0.124
4	Embedding -only	✓	✓	0.726 ±0.103	0.885 ±0.013	0.010 ±0.016	0.838 ±0.058	0.024 ±0.078
5	Embedding -only		✓	0.593 ±0.107	0.900 ±0.013	0.027 ±0.014	0.904 ±0.074	0.106 ±0.089
6	Embedding -only			0.243 ±0.123	0.888 ±0.012	0.070 ±0.015	0.751 ±0.150	0.209 ±0.104

Note: A check mark (✓) indicates that the corresponding component was enabled. Reg. denotes the combined use of L2 regularization, batch normalization, and dropout. Test AUC, AUC Gap, Test F_{pb} , and F_{pb} Gap are reported as mean ± standard deviation across 100 independent runs. AUC Gap and F_{pb} Gap are calculated as the corresponding training-set value minus the test-set value; smaller values indicate less overfitting. SHAP Spearman is reported as the mean ± SD of the pairwise Spearman rank correlations among SHAP-based feature-importance rankings from the 100 runs; higher values indicate greater ranking stability.

4. Discussion

4.1. A dual-mode framework for engineering-geological screening and assessment

To contextualize the operational value of the proposed model, its performance is compared with recent LSM studies in HK while accounting for differences in sample construction and prediction settings. Direct cross-study comparison of AUC values should be treated with caution because existing studies use different negative-sample strategies, predictor sets,

temporal settings, and validation designs. Negative samples have been constrained using landslide priors (Chen et al. 2024), mitigation information (Lyu et al. 2024), manual selection (Yu et al. 2025), or predefined spatial exclusion rules (Qiu et al. 2024). These procedures incorporate different amounts and types of prior information and can change the separability between landslide and non-landslide samples; therefore, their reported AUC values are not strictly comparable with those obtained under a fully automated whole-area sampling scheme. In a controlled comparison, the RF model achieved a test AUC of 0.8713 under global random sampling and 0.8953 after mitigation-informed non-landslide samples were added (Lyu et al. 2024), illustrating the sensitivity of model performance to sample construction. Dynamic yearly assessment and rainfall-driven spatiotemporal modeling also address prediction settings that differ from the present study (Ren et al. 2024; Ma et al. 2025) and are thus treated as contextual references rather than direct benchmarks. Under the fully automated sampling framework adopted here, the Embedding-only model achieved performance comparable to the global-random-sampling benchmark, for which an RF test AUC of 0.8713 was reported (Lyu et al. 2024). This result supports the operational viability of GeoFM embeddings as a competitive baseline for rapid automated mapping, rather than demonstrating strict superiority over studies based on more strongly curated samples.

Building upon this validation, this study proposes a dual-mode mapping framework (Fig. 6), using response urgency as the key criterion for selecting between two technical pipelines.

Level 1: Rapid Response Pipeline. This pipeline is engineered for emergency situations or data-scarce regions. Unlike traditional PN supervised learning, this mode leverages PU learning to overcome the critical bottleneck of negative sample unavailability during early

disaster stages. In traditional PN modeling, robust training requires both confirmed disaster points (positives) and the rigorous delineation of non-disaster zones (negatives) (Bekker and Davis 2020). The advantage of PU learning lies in its ability to train directly on reported disaster samples and unlabeled background data, eliminating the prerequisite for confirmed safety zones (Li et al. 2011). This algorithmic characteristic, combined with the low-preprocessing nature of GeoFM embeddings, creates a minimal-intervention automated system. Users need only provide disaster coordinates to generate an initial susceptibility map in a rapid and automated manner, bypassing time-consuming environmental data cleaning and negative labeling.

Level 2: Detailed Assessment Pipeline. This pipeline is tailored for scenarios with stringent requirements, such as post-disaster reconstruction planning or engineering site selection. It emphasizes the deep integration of multi-source data and logical consistency. In this workflow, GeoFM embeddings are combined with physical factors so that both surficial semantics and explicit physical controls are represented in the model. To prevent high-dimensional embedding features from overshadowing low-dimensional physical factors, a feature balancing mechanism is introduced via dimensionality reduction. Furthermore, the pipeline also incorporates interpretability analysis. By examining whether feature contributions are broadly consistent with physical understanding through SHAP analysis, the system provides iterative guidance for data screening, ensuring that the final susceptibility map is not only statistically accurate but also physically interpretable and robust.

The proposed framework should be viewed as a susceptibility screening tool rather than a substitute for site-specific geotechnical investigation. Its value lies in producing spatially

focused susceptibility information under limited-label conditions, which can support subsequent engineering-geological assessment and field verification.

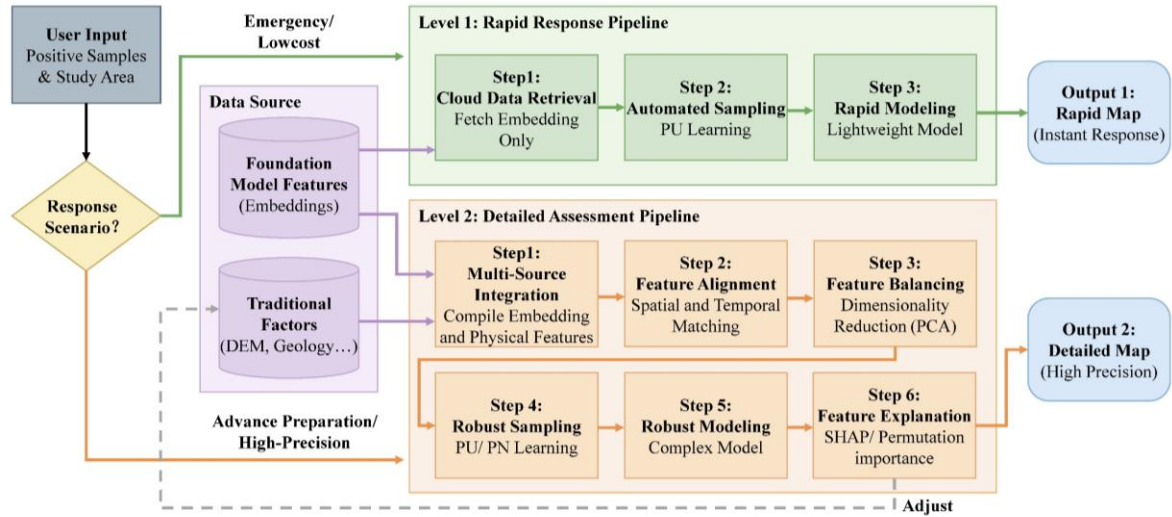


Fig. 6 Dual-mode framework for rapid engineering-geological screening and detailed landslide susceptibility assessment

4.2. GeoFM embeddings in local-scale LSM: roles, usage, and generalizability

4.2.1. Roles of GeoFM embeddings in local-scale LSM

The results suggest that the value of GeoFM embeddings in local-scale LSM lies not in replacing conventional conditioning factors, but in complementing them. In particular, the embeddings provide rich Earth-observation semantics that go well beyond the information content of simple satellite-derived variables such as NDVI and NDBI (Jakubik et al. 2023; Mai et al. 2024). Although categorical Land Use and Land Cover variables are still retained in the model, the embeddings offer a finer-grained and more continuous description of surface textures and material compositions, thereby capturing local environmental heterogeneity that discrete classification schemes may fail to resolve. As a result, when embeddings are fused

with explicit physical factors, the model can place less reliance on simplified surface-level proxies and assign greater importance to deeper controlling factors such as Lithology and TWI, which are more directly related to slope failure mechanisms (Guzzetti et al. 1999). This semantic-physical complementarity provides a plausible explanation for why the Fused model achieves both stronger predictive performance and more physically meaningful decision logic.

A more specific contribution of GeoFM embeddings can be observed in highly urbanized areas. In these areas, the Traditional-only model produces a terrain-driven false-positive pattern: artificial steep surfaces represented in the DTM, such as embankments and elevated roads, are assigned high susceptibility because their geometric characteristics resemble those of natural steep slopes. We refer to this pattern as “elevation deception”. This pattern is particularly evident in the Kowloon Peninsula, where the Traditional-only model predicts extensive high-susceptibility areas within the built environment (Fig. 4). By incorporating surface semantics beyond terrain geometry, the Fused model reduces these urban false positives while retaining high-susceptibility predictions in natural mountainous terrain. This comparison indicates that GeoFM embeddings complement explicit physical factors by refining the model’s representation of urban surface context, rather than replacing the geomorphological, geological, and hydrological controls of landslide susceptibility.

4.2.2. Temporal implications of GeoFM embeddings

Another important implication of GeoFM embeddings concerns their temporal character. Previous Hong Kong studies have incorporated temporal or InSAR information in different ways, including inventory augmentation (Chen et al. 2023), model enhancement and validation

(Chen et al. 2024; Yu et al. 2025), dynamic yearly assessment (Ma et al. 2025), and prior evaluation of InSAR applicability (Zhang et al. 2025). The AlphaEarth embeddings used here are generated from multi-temporal Earth observation data and therefore implicitly encode the spatio-temporal evolution of the land surface (Brown et al. 2025). This offers a streamlined way to represent cumulative environmental change, because the temporal dimension is already embedded in the feature representation rather than being introduced through a separate post-processing step. In operational terms, this property may reduce preprocessing complexity and enhance the practicality of rapid deployment.

However, the temporal advantage of embeddings should not be overstated. The present framework mainly exploits implicit temporal information to characterize the quasi-static condition of the surface environment, and this may not be sufficient to capture instantaneous triggering processes. In particular, the model is likely to remain less sensitive to short-term extreme rainfall, typhoon-driven disturbances, and other rapidly evolving triggers that are central to real-time hazard formation (Ghosh and Saha 2025). For this reason, the implicit temporal encoding of GeoFM should be viewed as a useful supplement to static susceptibility assessment, rather than as a complete replacement for explicitly dynamic warning variables. A promising next step would be to integrate real-time meteorological variables or dynamic rainfall sequences into the current framework, especially within the higher-precision assessment pipeline, so that quasi-static embedding features can be directly combined with short-term triggering information. Such an extension would help connect susceptibility-oriented mapping with more dynamic hazard assessment or early-warning applications.

4.2.3. Strategies for using high-dimensional GeoFM embeddings

The ablation results further indicate that the practical value of embeddings depends not only on whether they are used, but also on how they are used. From a purely predictive perspective, retaining raw high-dimensional embeddings may be acceptable when the sole objective is to maximize classification metrics, since deep neural networks are in principle capable of extracting non-linear patterns from complex inputs. Nevertheless, the results of this study suggest that raw-feature input comes at a substantial cost in interpretability and robustness. In the present dataset, the high-dimensional embeddings exhibited severe internal multicollinearity, a condition known to complicate model interpretation (Dormann et al. 2013). More generally, learned representations may encode multiple entangled factors of variation (Bengio et al. 2013). In our repeated experiments, directly combining the raw embedding block with a much smaller set of explicit physical variables was associated with less stable feature-importance rankings. We therefore interpret this instability as an empirical result of the current feature configuration, rather than as a universal consequence of high-dimensional representation learning. In this sense, the main benefit of dimensionality reduction and regularization is not simply noise reduction, but the preservation of a more stable and physically interpretable decision structure.

Accordingly, the results support a hierarchical recommendation for embedding deployment. In accuracy-first scenarios, especially where interpretability is not the primary concern, retaining raw embeddings remains a feasible option, as it may yield marginal gains in predictive metrics. By contrast, in hybrid modeling settings, particularly at the local scale, where embeddings are fused with explicit terrain, geology, and hydrology variables, a

dimensionality reduction strategy is strongly recommended. Compressing the embedding space helps construct a more balanced feature landscape, allowing the semantic contribution of GeoFM to be retained without overwhelming the physically interpretable variables. The resulting trade-off, namely a marginal sacrifice in raw accuracy in exchange for greater robustness, consistency, and interpretability, appears justified for practical landslide susceptibility assessment.

At the same time, the current feature-handling strategy still leaves room for methodological improvement. PCA is a classical and effective linear solution, but it may be insufficient for disentangling the complex non-linear structure of the embedding manifold (Hinton and Salakhutdinov 2006). Recent landslide studies have begun to explore more advanced deep learning strategies for robust feature extraction (Chen et al. 2024), and future work could therefore examine whether non-linear compression methods can make more effective use of GeoFM embeddings. In addition, the present study treats GeoFM as a frozen feature extractor and trains only the downstream classifier. While this design is computationally efficient and suitable for deployment, it may limit the model's sensitivity to subtle local geomorphological signatures. As fine-tuning becomes increasingly common in GeoFM-related studies, future research could explore unfreezing selected layers for domain adaptation so that the model can directly learn landslide-relevant visual signatures instead of relying only on general-purpose semantic features (Li et al. 2025).

4.2.4. Generalizability across regions and hazard types

Beyond the HK case, the broader appeal of GeoFM lies in its potential transferability. Because

foundation-model representations are designed to support adaptation across diverse Earth observation datasets and downstream tasks, they offer an attractive basis for rapid susceptibility assessment in data-scarce regions or newly affected areas. Existing evidence demonstrates transfer-learning gains across downstream Earth observation tasks, including land-cover classification and semantic segmentation (Cong et al. 2022), while related work has explored the use of AlphaEarth representations for burned-area mapping (Seydi 2025). These findings provide a rationale for testing the proposed rapid-response pipeline beyond HK.

That said, this potential remains largely prospective in the context of the present study. The empirical validation is currently confined to HK, and therefore the cross-regional robustness of the framework should not yet be assumed. Differences in climatic regimes, geological conditions, terrain morphology, and inventory completeness may all affect how well the learned semantic-physical relationships transfer to other regions. Similarly, although the general-purpose nature of GeoFM suggests possible utility for other natural hazards, such as floods and wildfires, this multi-hazard applicability still requires direct testing rather than conceptual extrapolation. Future work should therefore prioritize systematic validation across multiple regions and hazard types, so as to determine whether the present framework can evolve from a local landslide application into a more general emergency-response mapping tool.

5. Conclusions

This study investigates how GeoFM embeddings can be practically deployed in LSM under realistic label constraints. By integrating AlphaEarth embeddings with a PU-learning strategy

based on the PBLC algorithm and validating the workflow in HK, we draw the following conclusions:

(1) GeoFM embeddings provide a practical and competitive representation for PU-based local-scale LSM, while feature fusion yields the best overall mapping performance. Across the three feature configurations, the Embedding-only model provides a competitive and deployment-friendly baseline, whereas the Fused model consistently achieves the strongest predictive performance and delineates the most spatially concentrated high-susceptibility zones. In particular, the Fused model captures the majority of landslides within only 6.3% of the study area, achieving a landslide density 24 times higher than the background average. These findings further support a dual-mode application framework, in which Embedding-only PU models enable rapid susceptibility screening under time and data constraints, whereas Fused models are recommended for detailed assessment when higher predictive precision and more spatially selective delineation of high-susceptibility areas are required.

(2) Traditional physical factors and GeoFM embeddings play complementary roles in model reasoning. While physically interpretable factors remain central to susceptibility formation, high-dimensional embeddings contribute fine-grained environmental semantics that help resolve failure cases of purely physical logic. Notably, embeddings provide additional surface information that helps reduce false positives in highly urbanized areas affected by “elevation deception”. Moreover, their implicit spatio-temporal encoding may streamline the representation of cumulative environmental change, although it should be regarded as a supplement rather than a replacement for explicitly dynamic triggering variables. The results also suggest that GeoFM embeddings may partially substitute for some conventional satellite-

derived indices, such as NDVI and NDBI.

(3) Stable deployment of GeoFM embeddings requires targeted feature-space optimization and robustness checks. High-dimensional embeddings exhibit severe multicollinearity at local scales. Although directly using raw embeddings may bring marginal performance gains, it can lead to instability in feature-importance rankings and weaken interpretability. Dimensionality reduction and regularization are therefore recommended for robust deployment, especially when embeddings are fused with explicit physical variables and interpretability is an important concern.

Overall, the results indicate that GeoFM can be operationalized in local-scale LSM through a PU-learning workflow under realistic label constraints. The proposed framework is particularly useful for rapid engineering-geological screening, prioritization of field investigation, and detailed susceptibility assessment when both transferable embeddings and physical conditioning factors are available.

References

- Bekker J, Davis J (2020) Learning from positive and unlabeled data: a survey. *Mach Learn* 109:719–760. <https://doi.org/10.1007/s10994-020-05877-5>
- Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35:1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>
- Brown CF, Kazmierski MR, Pasquarella VJ et al (2025) AlphaEarth foundations: an

656 embedding field model for accurate and efficient global mapping from sparse label
657 data. arXiv preprint arXiv:2507.22291. <https://doi.org/10.48550/arXiv.2507.22291>

658 Chen H, Ma T, Shen L et al (2025a) Negative sample selection for landslide susceptibility
659 prediction: a hybrid optimization approach using an AHP-KDE multi-ring sampling
660 strategy. *Ecol Indic* 180:114304. <https://doi.org/10.1016/j.ecolind.2025.114304>

661 Chen L, Ma P, Fan X et al (2024) A knowledge-aware deep learning model for landslide
662 susceptibility assessment in Hong Kong. *Sci Total Environ* 941:173557.
663 <https://doi.org/10.1016/j.scitotenv.2024.173557>

664 Chen L, Ma P, Yu C et al (2023) Landslide susceptibility assessment in multiple urban slope
665 settings with a landslide inventory augmented by InSAR techniques. *Eng Geol*
666 327:107342. <https://doi.org/10.1016/j.enggeo.2023.107342>

667 Chen X, Li W, Hsu C-Y et al (2025b) Harnessing geospatial artificial intelligence and deep
668 learning for landslide inventory mapping: advances, challenges, and emerging
669 directions. *Remote Sens* 17:1856. <https://doi.org/10.3390/rs17111856>

670 Cong Y, Khanna S, Meng C et al (2022) SatMAE: pre-training transformers for temporal and
671 multi-spectral satellite imagery. In: *Advances in Neural Information Processing*
672 *Systems*. Curran Associates, Inc., pp 197–211. <https://doi.org/10.52202/068431-0015>

673 Dormann CF, Elith J, Bacher S et al (2013) Collinearity: a review of methods to deal with it
674 and a simulation study evaluating their performance. *Ecography* 36:27–46.
675 <https://doi.org/10.1111/j.1600-0587.2012.07348.x>

676 Dou H, He J, Huang S et al (2023) Influences of non-landslide sample selection strategies on
677 landslide susceptibility mapping by machine learning. *Geomat Nat Hazards Risk*
678 14:2285719. <https://doi.org/10.1080/19475705.2023.2285719>

679 Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. *Nat*
680 *Hazards Earth Syst Sci* 18:2161–2181. <https://doi.org/10.5194/nhess-18-2161-2018>

681 Fu Z, Ma H, Wang F et al (2024) An integrated framework of positive-unlabeled and
682 imbalanced learning for landslide susceptibility mapping. *IEEE J Sel Top Appl Earth*
683 *Obs Remote Sens* 17:15596–15611. <https://doi.org/10.1109/JSTARS.2024.3452182>

684 Gao L, Zhang LM, Chen HX et al (2021) Topography and geology effects on travel distances
685 of natural terrain landslides: Evidence from a large multi-temporal landslide inventory
686 in Hong Kong. *Eng Geol* 292:106266. <https://doi.org/10.1016/j.enggeo.2021.106266>

687 Ghosh K, Saha R (2025) A critical review on rainfall induced landslide occurrences and early
688 warning systems. *Geomech Geoengin* 20:1052–1074.
689 <https://doi.org/10.1080/17486025.2025.2496302>

690 Guo X, Lao J, Dang B et al (2024) SkySense: a multi-modal remote sensing foundation
691 model towards universal interpretation for Earth observation imagery. In: 2024
692 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp
693 27662–27673. <https://doi.org/10.1109/CVPR52733.2024.02613>

694 Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a
695 review of current techniques and their application in a multi-scale study, Central Italy.

696 Geomorphology 31:181–216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)

697 Haque U, Blum P, da Silva PF et al (2016) Fatal landslides in Europe. *Landslides* 13:1545–
698 1554. <https://doi.org/10.1007/s10346-016-0689-3>

699 Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural
700 networks. *Science* 313:504–507. <https://doi.org/10.1126/science.1127647>

701 Hong D, Zhang B, Li X et al (2024) SpectralGPT: spectral remote sensing foundation model.
702 *IEEE Trans Pattern Anal Mach Intell* 46:5227–5244.
703 <https://doi.org/10.1109/TPAMI.2024.3362475>

704 Jakubik J, Roy S, Phillips CE et al (2023) Foundation models for generalist geospatial
705 artificial intelligence. *arXiv preprint arXiv:2310.18660*.
706 <https://doi.org/10.48550/arXiv.2310.18660>

707 Jakubik J, Yang F, Blumenstiel B et al (2025) TerraMind: large-scale generative
708 multimodality for Earth observation. In: *Proceedings of the IEEE/CVF International*
709 *Conference on Computer Vision (ICCV)*. pp 7383–7394.
710 <https://doi.org/10.1109/ICCV51701.2025.00693>

711 Jiang Y, Wang W, Zou L et al (2025) Investigating landslide data balancing for susceptibility
712 mapping using generative and machine learning models. *Landslides* 22:189–204.
713 <https://doi.org/10.1007/s10346-024-02352-3>

714 Khaire UM, Dhanalakshmi R (2022) Stability of feature selection algorithm: a review. *J King*

715 Saud Univ - Comput Inf Sci 34:1060–1073.
 716 <https://doi.org/10.1016/j.jksuci.2019.06.012>

717 Kim H, Oikonomou O (2025) ZeroFlood: Flood Hazard Mapping from Single-Modality SAR
 718 Using Geo-Foundation Models. arXiv preprint arXiv:2510.23364.
 719 <https://doi.org/10.48550/arXiv.2510.23364>

720 Ko FWY, Lo FLC (2016) Rainfall-based landslide susceptibility analysis for natural terrain in
 721 Hong Kong - A direct stock-taking approach. Eng Geol 215:95–107.
 722 <https://doi.org/10.1016/j.enggeo.2016.11.001>

723 Kraabel N, Liu J, Bian Y et al (2025) StefaLand: an efficient geoscience foundation model
 724 that improves dynamic land-surface predictions. arXiv preprint arXiv:2509.17942.
 725 <https://doi.org/10.48550/arXiv.2509.17942>

726 Li W, Guo Q (2021) Plotting receiver operating characteristic and precision-recall curves
 727 from presence and background data. Ecol Evol 11:10192–10206.
 728 <https://doi.org/10.1002/ece3.7826>

729 Li W, Guo Q, Elkan C (2011) Can we model the probability of presence of species without
 730 absence data? Ecography 34:1096–1105. [https://doi.org/10.1111/j.1600-](https://doi.org/10.1111/j.1600-0587.2011.06888.x)
 731 [0587.2011.06888.x](https://doi.org/10.1111/j.1600-0587.2011.06888.x)

732 Li W, Guo Q, Elkan C (2021) One-class remote sensing classification from positive and
 733 unlabeled background data. IEEE J Sel Top Appl Earth Obs Remote Sens 14:730–
 734 746. <https://doi.org/10.1109/JSTARS.2020.3025451>

735 Li W, Liu Y, Liu Z et al (2022) A positive-unlabeled learning algorithm for urban flood
736 susceptibility modeling. *Land* 11:1971. <https://doi.org/10.3390/land11111971>

737 Li W, Wang S, Lee H et al (2025) Landslide hazard mapping with geospatial foundation
738 models: geographical generalizability, data scarcity, and band adaptability. *arXiv*
739 preprint *arXiv:2511.04474*. <https://doi.org/10.48550/arXiv.2511.04474>

740 Lin T-Y, Goyal P, Girshick R et al (2017) Focal loss for dense object detection. In: 2017 IEEE
741 International Conference on Computer Vision (ICCV). pp 2999–3007.
742 <https://doi.org/10.1109/ICCV.2017.324>

743 Liu R, Yang X, Xu C et al (2022) Comparative study of convolutional neural network and
744 conventional machine learning methods for landslide susceptibility mapping. *Remote*
745 *Sens* 14:321. <https://doi.org/10.3390/rs14020321>

746 Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In:
747 *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp
748 4765–4774

749 Lyu H-M, Yin Z-Y, Hicher P-Y, Laouafa F (2024) Incorporating mitigation strategies in
750 machine learning for landslide susceptibility prediction. *Geosci Front* 15:101869.
751 <https://doi.org/10.1016/j.gsf.2024.101869>

752 Ma P, Chen L, Yu C et al (2025) Dynamic landslide susceptibility mapping over last three
753 decades to uncover variations in landslide causation in subtropical urban mountainous
754 areas. *Remote Sens Environ* 326:114800. <https://doi.org/10.1016/j.rse.2025.114800>

755 Mai G, Huang W, Sun J et al (2024) On the opportunities and challenges of foundation
 756 models for GeoAI (vision paper). *ACM Trans Spat Algorithms Syst* 10:11:1-11:46.
 757 <https://doi.org/10.1145/3653070>

758 Meena SR, Puliero S, Bhuyan K et al (2022) Assessing the importance of conditioning factor
 759 selection in landslide susceptibility for the province of Belluno (region of Veneto,
 760 northeastern Italy). *Nat Hazards Earth Syst Sci* 22:1395–1417.
 761 <https://doi.org/10.5194/nhess-22-1395-2022>

762 Merghadi A, Yunus AP, Dou J et al (2020) Machine learning methods for landslide
 763 susceptibility studies: a comparative overview of algorithm performance. *Earth-Sci*
 764 *Rev* 207:103225. <https://doi.org/10.1016/j.earscirev.2020.103225>

765 Qiu A, Wang Q, Chen Y et al (2024) Landslide susceptibility assessment in Hong Kong with
 766 consideration of spatio-temporal consistency. *Appl Sci* 14:10654.
 767 <https://doi.org/10.3390/app142210654>

768 Reed CJ, Gupta R, Li S et al (2023) Scale-MAE: a scale-aware masked autoencoder for
 769 multiscale geospatial representation learning. In: *Proceedings of the IEEE/CVF*
 770 *International Conference on Computer Vision (ICCV)*. pp 4065–4076.
 771 <https://doi.org/10.1109/ICCV51070.2023.00378>

772 Reichenbach P, Rossi M, Malamud BD et al (2018) A review of statistically-based landslide
 773 susceptibility models. *Earth-Sci Rev* 180:60–91.
 774 <https://doi.org/10.1016/j.earscirev.2018.03.001>

775 Ren T, Gao L, Gong W (2024) An ensemble of dynamic rainfall index and machine learning
 776 method for spatiotemporal landslide susceptibility modeling. *Landslides* 21:257–273.
 777 <https://doi.org/10.1007/s10346-023-02152-1>

778 Seydi ST (2025) Deep learning-based burned area mapping using Bi-temporal siamese
 779 networks and AlphaEarth foundation datasets. *arXiv preprint arXiv:2509.07852*.
 780 <https://doi.org/10.48550/arXiv.2509.07852>

781 Stow M, Stewart AA (2025) Empirical analysis of SHAP stability under data corruption
 782 across datasets and model architectures. *Int Adv Res J Sci Eng Technol* 12:92–110.
 783 <https://doi.org/10.17148/IARJSET.2025.12810>

784 Szwarcman D, Roy S, Fraccaro P et al (2026) Prithvi-EO-2.0: a versatile multitemporal
 785 foundation model for Earth observation applications. *IEEE Trans Geosci Remote Sens*
 786 64:1–20. <https://doi.org/10.1109/TGRS.2025.3642610>

787 Wang C, Lin Q, Wang L et al (2022) The influences of the spatial extent selection for non-
 788 landslide samples on statistical-based landslide susceptibility modelling: a case study
 789 of Anhui Province in China. *Nat Hazards* 112:1967–1988.
 790 <https://doi.org/10.1007/s11069-022-05252-8>

791 Wong HN (2009) Rising to the challenges of natural terrain landslides. In: *Natural Hillside:
 792 Study and Risk Management Measures, Proceedings of the 29th Annual Seminar of
 793 the HKIE Geotechnical Division. The Hong Kong Institution of Engineers, Hong
 794 Kong, pp 15–53*

- Wu B, Qiu W, Jia J, Liu N (2021) Landslide susceptibility modeling using bagging-based positive-unlabeled learning. *IEEE Geosci Remote Sens Lett* 18:766–770.
<https://doi.org/10.1109/LGRS.2020.2989497>
- Yao J, Qin S, Qiao S et al (2022) Application of a two-step sampling strategy based on deep neural network for landslide susceptibility mapping. *Bull Eng Geol Environ* 81:148.
<https://doi.org/10.1007/s10064-022-02615-0>
- Yu M, Chen L, Fan X et al (2025) Validating and enhancing data-driven landslide susceptibility prediction by model explanation and MT-InSAR techniques. *Int J Digit Earth* 18:2509857. <https://doi.org/10.1080/17538947.2025.2509857>
- Zhang D, Jindal D, Roy N et al (2024) Enhancing landslide susceptibility mapping using a positive-unlabeled machine learning approach: a case study in Chamoli, India. *Geoenvironmental Disasters* 11:21. <https://doi.org/10.1186/s40677-024-00281-w>
- Zhang Z, Wu S, Zhao C et al (2025) Evaluation of InSAR applicability using a new multi-index and optical imagery: A case study in the Guangdong-Hong Kong-Macao greater bay area, China. *Remote Sens Appl Soc Environ* 37:101474.
<https://doi.org/10.1016/j.rsase.2025.101474>

Statements and Declarations

Funding

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2022B1515130001 and 2023A1515011369).

Competing Interests

The authors declare that they have no financial interests.

Author Contributions

Mingxuan Li: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. **Wenkai Li:** Conceptualization, Methodology, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Hongliang Liu:** Formal analysis, Validation, Writing – review & editing. **Changshan Liang:** Formal analysis, Validation, Writing – review & editing.

Data Availability

Historical landslide data were collected from the Enhanced Natural Terrain Landslide Inventory (ENTLI), published by the Civil Engineering and Development Department of Hong Kong. The underlying remote-sensing and ancillary datasets were obtained from the sources listed in Table 1 and remain subject to the terms and licenses of their respective providers.