



1 A Study of Small-scale Landslides Susceptibility Dynamic 2 Assessment in South Chian: Merging SBAS-InSAR and 3 Machine Learning

4 Zhu liang^{1,2,3}, Jingxin Hou^{1,2,3}, Yang Liu^{1,2,3,*}, Ting Wang⁴, Guochao Liu^{1,2,3}, Chunshuai Si^{1,2,3}
5 and Jun Wu^{1,2,3}

6 ¹Guangzhou Urban Planning & Design Survey Research Institute Co., Ltd, Guangzhou, 510060, China

7 ²Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou,
8 Guangzhou, 510060, China

9 ³Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou, 510060,
10 China

11 ⁴Guangzhou Meteorological Observatory, Guangzhou, 511430

12 * Correspondence: liuyang@gzpi.com.cn

13 **Abstract.** Shallow landslides, triggered by factors such as heavy rainfall and human engineering activities, are
14 characterized by their sudden occurrence, wide distribution, and small scale, which pose significant threats to
15 human life, especially in southern China. Landslide susceptibility assessment (LSA) is crucial for disaster
16 prevention, mitigation, and land-use planning. Traditional assessment methods, such as field surveys and
17 statistical models, predominantly depend on static geological environment factors, which exhibit inherent
18 limitations in capturing spatiotemporal information on dynamic surface deformation. This deficiency directly
19 leads to a lack of timeliness in hazard assessment, which is particularly important for sudden disasters. In this
20 study, small Baseline Subset (SBAS - InSAR) and machine learning were married to explore a dynamic method
21 in LSA, and the Conghua district in Guangzhou was selected as the study area. First, a dataset consisting of 326
22 historical landslides and 10 static factors, and 2 dynamic factors in the area was prepared. Then, the dataset was
23 divided into two parts, one for modeling training (80%) and the other for testing (20%). Third, factors and samples
24 were involved in the modeling of Random Forest (RF), Light Gradient Boosting Machine (LightGBM), and
25 Extreme Gradient Boosting (XGBoost). Finally, the performance of these models was validated through the area
26 under the curve (AUC) and compared with the models analyzed by 10 static factors only. The results show that
27 the XGBoost model exhibits the best performance, with an AUC of 0.933, and the models considered dynamic
28 factors all performed better than that of static factors only. This study demonstrates that merging SBAS-InSAR
29 and machine learning can provide a reliable technical approach for dynamic assessment of small-scale LSA, which
30 is of great significance for targeted disaster prevention and mitigation in rural and hilly areas of South China.

31 **Keywords:** Landslide susceptibility map; Small-scale; SBAS – InSAR; Machine learning; Dynamic evaluation

32 1. Introduction

33 Small-scale shallow landslides are a common type of geological disaster, usually occurring in the upper soil layer
34 or weathered rock layer (Ado et al.,2022; Castelli et al.,2025; Sun et al.,2025). Under the combined influence of natural
35 factors (e.g., extreme rainfall, geological weathering) and human activities (e.g., rural road construction, farmland
36 expansion), small-scale landslides have become one of the most frequent geological disasters in southern China
37 (Zhang et al., 2021; Wang et al., 2021). Unlike large-scale landslides, small-scale landslides have small volumes



38 (usually 10 - 1000 m³) and shallow sliding depths (1 - 3 m), but they are widely distributed, highly sudden, and
39 difficult to monitor (Cui et al.,2023). In recent years, under the influence of global climate change, extreme rainfall
40 events have occurred frequently, and the intensity and frequency of shallow landslides have increased significantly.
41 With the rapid development of urbanization and infrastructure construction, human engineering activities further
42 increase the risk of shallow landslides.

43 Traditional LSA methods mainly include field survey methods, remote sensing image interpretation methods,
44 and statistical model methods. Field survey methods are time-consuming, labor-intensive, and have limited spatial
45 coverage, making it difficult to carry out large-area assessments (Kovács et al.,2019). Statistical model methods,
46 such as the Logistic Regression model, clustering analysis, require a large number of historical landslide data and
47 cannot handle the nonlinear relationship between factors (Huang et al.,2024; Nwazelibé et al.,2024). In recent years,
48 machine learning methods have been widely used in LSA due to their strong ability to handle nonlinear and high-
49 dimensional data (Huang et al.,2024; Zhou et al.,2025). However, most of these studies use static influencing factors
50 (e.g., lithology, slope, rainfall) and ignore the dynamic changes in slope stability, such as subtle surface
51 deformation before landslide initiation. Small-scale landslides often occur after a period of accelerated surface
52 deformation (e.g., deformation rate increases from -5 mm/year to -15 mm/year within 1 - 2 months before sliding),
53 which is a key precursor of landslide occurrence (Wang et al.,2024).

54 Traditional remote sensing methods (e.g., optical remote sensing) can only identify existing landslides but
55 cannot capture subtle surface deformation. Field monitoring methods (e.g., inclinometers, GNSS) can measure
56 surface deformation with high accuracy, but they are limited by high cost and low spatial coverage, making it
57 difficult to apply in large - area small - scale landslide assessment (Ebrahim et al.,2024; Nie et al.,2025; Wang et
58 al.,2024). Therefore, the lack of dynamic deformation information is the main bottleneck restricting the accuracy
59 of small-scale landslide susceptibility assessment. Compared with other remote sensing methods (e.g., PS -
60 InSAR), SBAS - InSAR has better adaptability to vegetated areas (which account for over 60% of the study area
61 in South China) by using small baseline interferometric pairs and extracting deformation information from
62 distributed scatterers (e.g., vegetation, soil) (Hussain et al.,2025; Zhang et al.,2025; Guo et al.,2024).

63 The study is to explore and compare the performance of merging SBAS-InSAR technology with different
64 machine learning models (RF, LightGBM, and XGBoost) in the assessment of small-scale landslide susceptibility.
65 Conghua dissect, Guangzhou was selected as the study area.

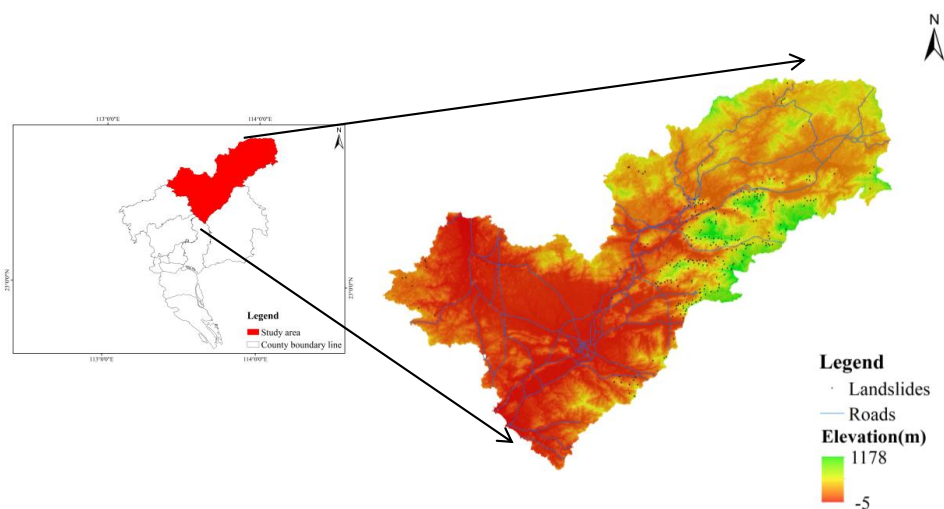
66 2. Study Area and Data Sources

67 2.1 Overview of the Study Area

68 The study area is located in the mountainous area of Guangdong Province, China, with a geographical coordinate
69 range of 109°20' - 110°00'E and 28°10' - 28°50'N, covering an area of approximately 2,500 km²(Fig.1). The study
70 area has a subtropical monsoon climate, with an average annual rainfall of 1,200 - 1,600 mm, and the rainfall is
71 mainly concentrated in the flood season (May - September), accounting for more than 70% of the annual rainfall.
72 The terrain in the study area is complex and undulating, with an elevation range of 200 - 1,800 m. The main
73 landforms are mountains, hills, and valleys. The stratum in the study area is mainly composed of Devonian
74 sandstone, shale, Carboniferous limestone, and Quaternary loose deposits. The geological structure is relatively
75 complex, with faults and folds developed. The study area is an area with frequent shallow landslide disasters. In



76 recent years, many shallow landslides have occurred, mainly triggered by heavy rainfall, house and road
77 construction. The number of landslides recorded in the study area is 361, with the volume ranging from 15 m³ to
78 400 m³(Fig.2). These landslides have caused serious damage to local infrastructure and people's lives and property,
79 so it is urgent to carry out a shallow landslide susceptibility assessment in this area.



80

81 **Figure 1: Location of the study area showing elevation and landslide samples.**



82



83
84 **Figure 2: Photos of the on-site investigation of shallow landslides.**

85 **2.2 Data Sources and Preprocessing**

86 **2.2.1 Sentinel - 1A/B SAR Data**

87 The SAR data used in this study are Sentinel - 1A/B C - band SAR images provided by the European Space
88 Agency (ESA). The data acquisition period is from January 2024 to December 2025, with a total of 48 images.
89 The imaging mode is Interferometric Wide (IW) mode, with a spatial resolution of 5 m (azimuth) × 20 m (range)
90 and a temporal baseline of 12 days. The main parameters of the Sentinel - 1A/B SAR data are shown in Table 1.

91 **Table 1. Main Parameters of Sentinel - 1A/B SAR Data**

Parameter	Value
Satellite	Sentinel - 1A/B
Sensor Band	C - band (5.405 GHz)
Imaging Mode	IW
Spatial Resolution	Azimuth: 5 m; Range: 20 m
Swath Width	250 km
Temporal Baseline	12 days
Polarization	VV + VH
Acquisition Period	January 2024- December 2025
Number of Images	48

92 The preprocessing of Sentinel - 1A/B SAR data is carried out using the Sentinel Application Platform (SNAP)
93 software.



94 **2.2.2 Digital Elevation Model (DEM) Data**

95 The DEM data used in this study were collected and interpreted independently by the research institution. with a
96 spatial resolution of 10 m. The DEM data are mainly used for geometric correction of SAR images, generation of
97 topographic phase in interferograms, and extraction of topographic factors (slope, aspect, elevation, curvature) for
98 shallow landslide susceptibility assessment.

99 **2.2.3 Geological Data**

100 The geological data used in this study include the 1:50,000 geological map of the study area and the lithology and
101 geological structure reports provided by the local geological survey department. The lithology data are classified
102 into five categories: sandstone, shale, limestone, Quaternary loose deposits, and others. The geological structure
103 data mainly include the distribution of faults and folds. These geological data are digitized using ArcGIS software
104 to form vector data, which are used as important evaluation indices for shallow landslide susceptibility assessment.

105 **2.2.4 Hydrological Data**

106 The hydrological data used in this study include the river distribution data. The river distribution data are obtained
107 from the 1:100,000 topographic map, and the distance from each mapping unit to the nearest river is calculated
108 using ArcGIS software. These hydrological data are used to reflect the influence of rainfall and surface water on
109 shallow landslide occurrence.

110 **2.2.5 Human Activity Data**

111 The human activity data used in this study include the road distribution data. The road distribution data are
112 obtained from the 1:100,000 topographic map, and the distance from each mapping unit to the nearest road is
113 calculated using ArcGIS software. These human activity data are used to reflect the influence of human
114 engineering activities on shallow landslide occurrence.

115 **2.2.6 Normalized Differential Vegetation Index (NDVI)**

116 The vegetation coverage used in this study includes NDVI, which was obtained from the MODIS satellite images
117 provided by NASA.

118 **Table 2. Summary of data information and functions**

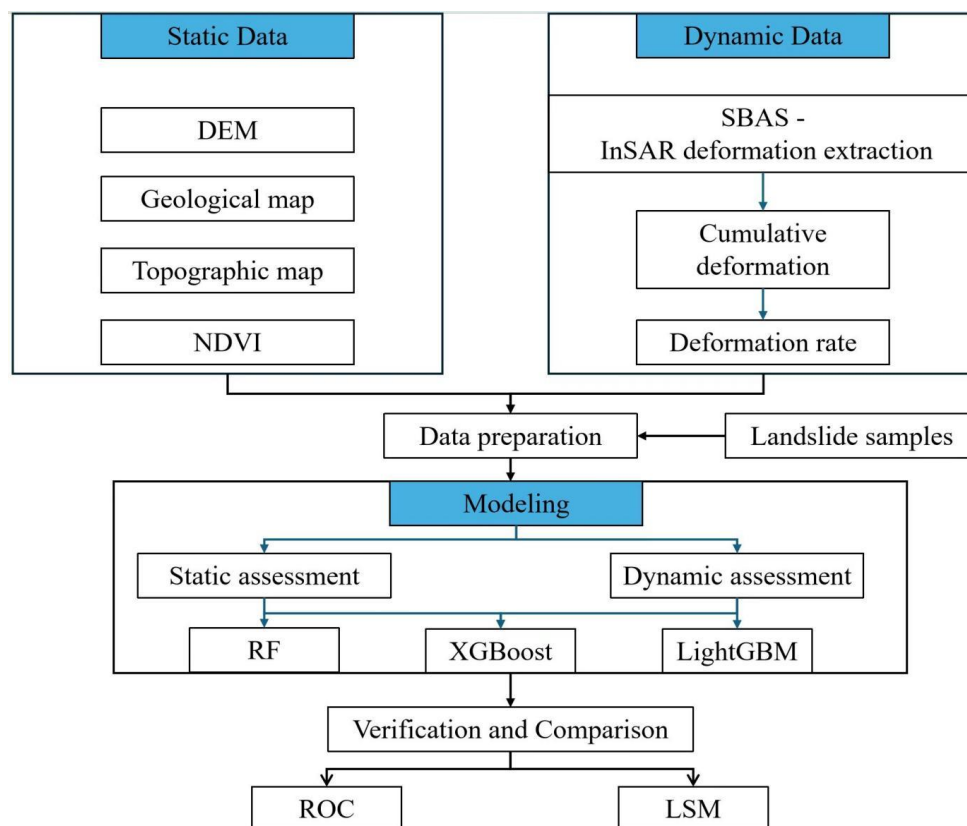
Items	Scale	Function	Source
Sentinel - 1A/B	5 m× 20 m	Extracting surface	European Space
SAR Data		deformation	Agency (ESA)
DEM	10 m× 10 m	Extraction of	Interpreted
		topographic factors	independently
Geological map	1:50,000	Extraction of	Geological Survey
		lithology and faults	Department
Topographic map	1:100,000	Extraction of river	Geological Survey
		and road distribution	Department
NDVI		Reflecting	NASA



		Vegetation coverage	
Landslide samples	15 m ³ to 400 m ³	Model training and validation	Survey reports, field surveys, and remote sensing image interpretation

119 **3. Methods**

120 The research framework of this study is shown in Fig.3, which mainly includes 4 parts: SBAS - InSAR surface
 121 deformation extraction, data preparation, machine learning modeling, verification, and comparison.



122

123 **Figure 3: Workflow of the study.**

124 **3.1 SBAS - InSAR Method for Surface Deformation Extraction**

125 The Small Baseline Subset (SBAS - InSAR) method is a Time-Series InSAR technique proposed by Berardino et
 126 al. (2002). This method selects SAR image pairs characterized by small spatial and temporal baselines, so as to
 127 form interferometric pairs. This approach can effectively mitigate the effects of temporal and spatial decorrelation
 128 as well as atmospheric delay, thereby enhancing the accuracy of surface deformation measurement:



129
$$\frac{N+1}{2} \leq M \leq \frac{N(N+1)}{2} \quad (1)$$

130 Where N is SAR images, M is interferograms.

131 Taking t_0 as the reference moment, $\varphi(t_i)$ denotes the differential phase at time t_i (where $i = 1, 2, \dots, N$). For
132 the m -th differential interferogram generated between time instants t_a and t_b (with $t_a > t_b$), the interference phase
133 value corresponding to any pixel (x, r) is:

134
$$\delta\varphi_m(x, r) = \varphi(t_a, x, r) - \varphi(t_b, x, r) \approx \frac{4\pi}{\lambda} [d(t_a, x, r) - d(t_b, x, r)] \quad (2)$$

135 Where $\delta\varphi_m(x, r)$ denotes the interferometric phase of pixel (x, r) in the m -th interferogram; λ represents the
136 wavelength of the satellite-borne radar; and $d(t_a, x, r)$ as well as $d(t_b, x, r)$ respectively stand for the cumulative
137 line-of-sight (LOS) deformations at times t_a and t_b , with both deformations referenced to the reference time t_0 .

138 In accordance with the fundamental principles of the Small Baseline Subset Interferometric Synthetic
139 Aperture Radar (SBAS-InSAR) technique, its standard workflow comprises six core stages, which are specifically
140 defined as follows: data preprocessing, baseline network construction, interferogram generation, phase
141 unwrapping, surface deformation estimation, and geocoding:

142
$$vm = \frac{\varphi_m - \varphi_{m-1}}{t_m - t_{m-1}} \quad (3)$$

143 Assuming that the deformation rate between two adjacent interferograms is denoted as v_k , k and (where k and
144 $k-1$ represent the serial numbers of SAR images, respectively). the value of the m -th differential interferogram
145 can be reformulated as follows:

146
$$\delta\varphi_m = \sum_{k=t, m+1}^{t_A, m} (t_k - t_{k-1}) v_k \quad (4)$$

147 In other words, this value represents the integral of the deformation rate v_k over each time period
148 corresponding to the time intervals between the master and slave SAR images. On this basis, Equation (4) can be
149 rewritten into the following matrix form:

150
$$Bv = \delta\varphi \quad (5)$$

151 Equation (5) corresponds to a matrix (denoted as B) with a dimension of $M \times N$. When there are multiple
152 small baseline sets involved, the matrix B exhibits rank deficiency. To address this issue, the Singular Value
153 Decomposition (SVD) method can be employed to decompose the coefficient matrix B—a process that facilitates
154 the joint inversion of multiple small baseline sets. Accordingly, a least squares solution that achieves the
155 minimization of the norm associated with the cumulative deformation variable can be derived. By virtue of this
156 solution, the accurate estimation of the deformation variable is rendered feasible.

157 3.2 Data preparation

158 3.2.1 Landslide Inventory

159 The historical landslide data used in this study are mainly obtained from the local geological disaster survey
160 reports, field surveys, and remote sensing image interpretation. A total of 361 historical shallow landslide points
161 were collected (as shown in Fig.1), including the location, scale, and occurrence time of the landslides. These
162 samples are used for model training (70%) and validation (30%).

163 3.2.2 Conditioning factors and mapping unit

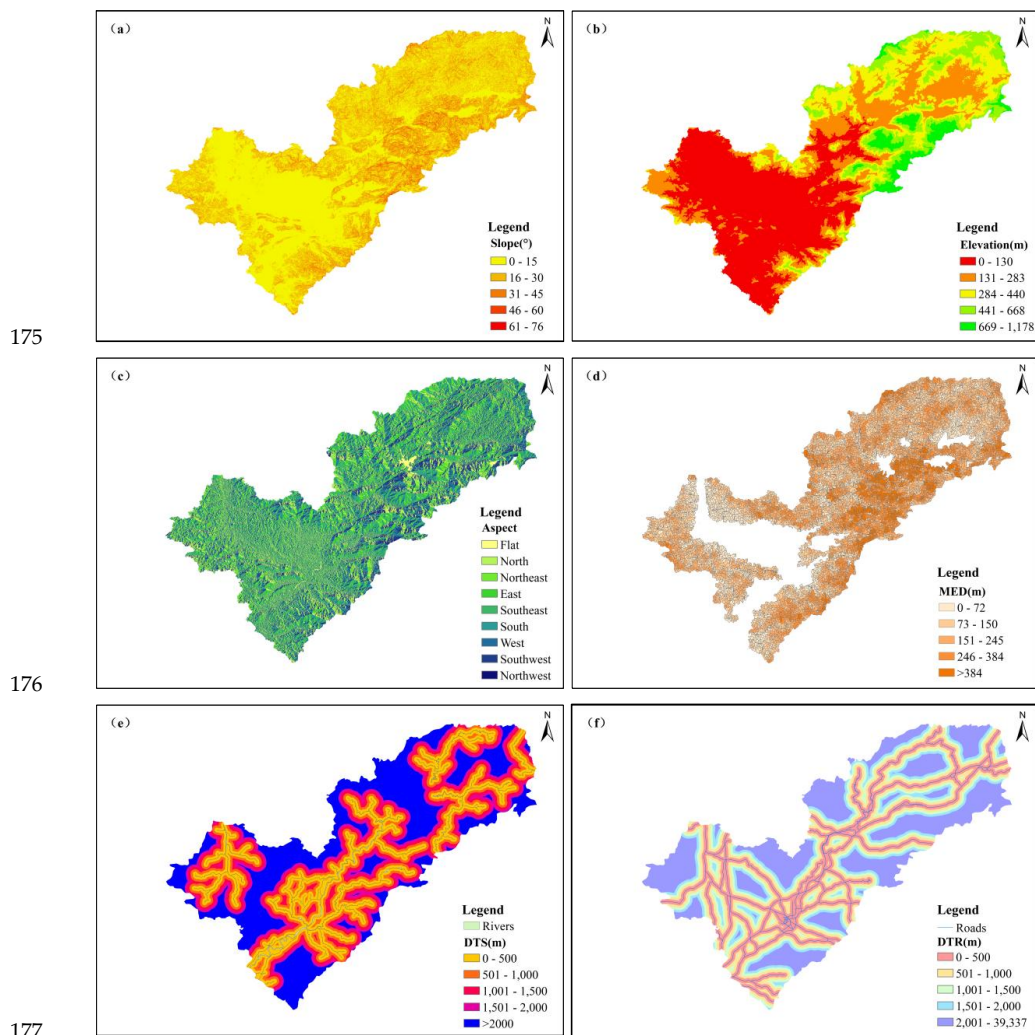
164 The conditioning factors selected in the study consisted of static and dynamic factors. Static factors included slope,



165 aspect, elevation, curvature, distance to faults, distance to rivers, distance to roads, and NDVI (Fig.4). Dynamic
166 factors included cumulative deformation and deformation rate (Fig.5). The specific indexes and their classification
167 standards are shown in Table 3.

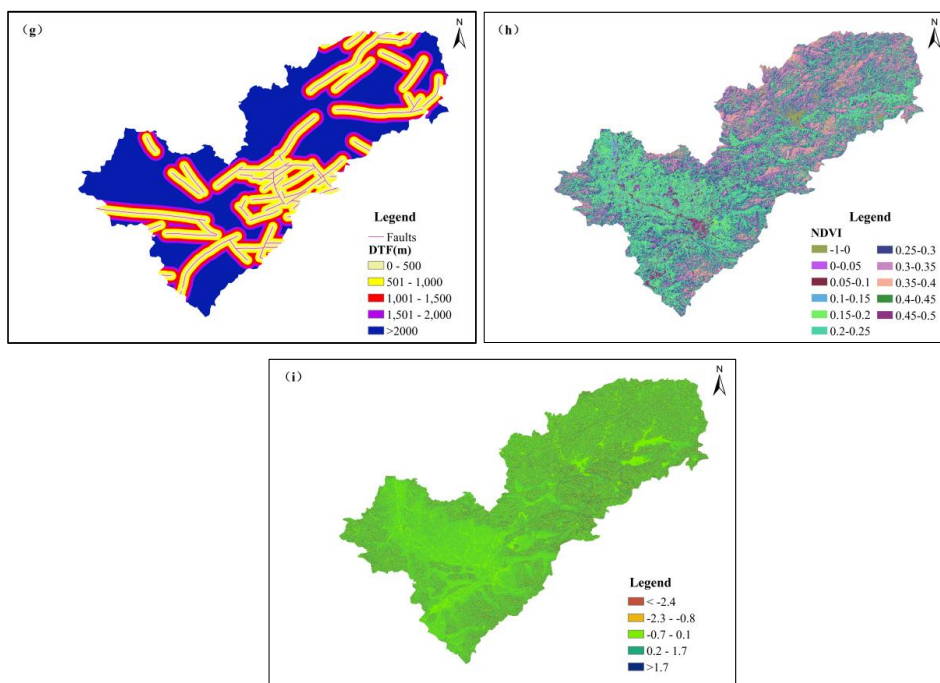
168 Prior to LSA, a suitable mapping unit must first be determined. Grid cells and slope units are two of the most
169 widely used mapping units, and a detailed comparison of the two can be found in other relevant literature (Lie et
170 al.,2024; Liu et al.,2024). In terms of preserving topographic conditions and distinguishing landslide locations, slope
171 units exhibit superior performance. Accordingly, slope units were adopted as the mapping unit, and the study area
172 was divided into 10426 units.

173 The study utilized ArcGIS 10.7 for the processing of thematic map generation, unit division, and data analysis,
174 including reclassification, distance analysis, and regional analysis etc.





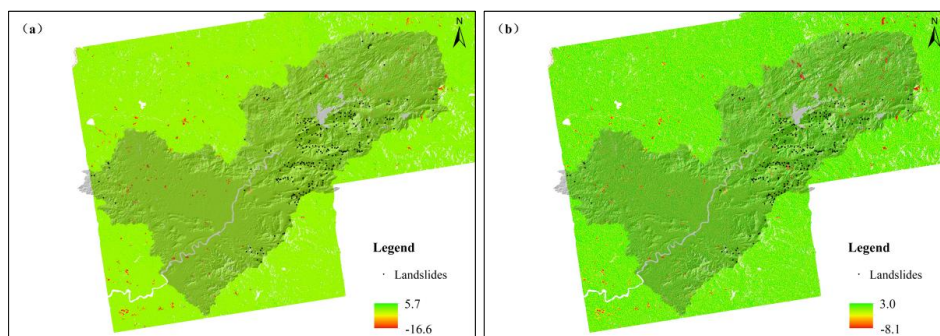
178



179

180 **Figure 4: Thematic map of static factors:(a) Slope; (b) Elevation; (c) Aspect; (d) Maximum elevation difference; (e)**
 181 **distance to rivers; (f) distance to roads; (g) distance to faults; (h) NDVI; (i) Profile curvature.**

182



183

Figure 5: Thematic map of dynamic factors: (a) Deformation; (b) Deformation rate.

184

Table 3. Shallow landslide susceptibility conditioning factors and classification standards

Category	Conditioning factors	Zone
Static factors	Slope (°)	0~15
		15~30
		30~45
		45~60
		>60
	Aspect	Flat
		North



		East
		Northeast
		Southeast
		South
		Southwest
		West
	Elevation(m)	0~228.4
		228.5~465.8
		465.9~703.2
		703.3~940.6
		>940.6
	Distance to faults(m)	0~500
		500~1000
		1000~1500
		1500~2000
		>2000
	Distance to rivers(m)	0~500
		500~1000
		1000~1500
		1500~2000
		>2000
	Distance to roads(m)	0~500
		500~1000
		1000~1500
		1500~2000
		>2000
	Profile curvature	<-2.4
		-2.3~-0.8
		-0.7~1
		0.1~1.7
		>1.7
	Maximum elevation difference(m)	0~157.8
157.9~315.6		
315.7~473.4		
473.5~631.2		
>631.3		
NDVI	-1~0.5	
Dynamic factors	Cumulative deformation (cm)	-16.6~5.7



	Deformation rate (cm/y)	-8.1~3.0
--	----------------------------	----------

185 **3.3 Modeling**

186 **3.3.1 RF**

187 The Random Forest (RF) algorithm, an ensemble learning method composed of multiple tree-based classifiers,
188 has gained widespread recognition and application in LSA (Li et al.,2024) . During the model training phase, K
189 sample subsets are generated from the original training dataset through the bootstrap sampling technique. The
190 samples that are not included in these subsets, known as out - of - bag (OOB) data, serve as a valuable resource
191 for internal error assessment of the model. Each of these K sample subsets is used to train an individual decision
192 tree, which acts as a base learner, resulting in K distinct classification outcomes (Breiman et al.,2021). For the final
193 classification result, a voting mechanism is adopted instead of a simple weighting approach—each decision tree's
194 classification result is treated as a vote, and the category that receives the majority of votes is determined as the
195 ultimate prediction of the RF model. A key advantage of the RF algorithm is its inherent ability to mitigate over-
196 fitting.

197 **3.3.2 LGBM**

198 The Light Gradient Boosting Machine (LGBM) stands out as a high-efficiency, tree-structured gradient boosting
199 algorithm (Ke et al.,2017). Distinguishing itself from conventional approaches, it adopts a leaf-wise tree growth
200 mechanism with restricted depth, replacing the traditional level-wise strategy—an innovation that simultaneously
201 mitigates overfitting and enhances predictive accuracy. The core of the leaf-wise method lies in prioritizing the
202 leaf node with the maximum gain for splitting at each iteration. In comparison to the level-wise strategy, which
203 expands all nodes at a given depth uniformly, the leaf-wise approach minimizes more of the errors associated with
204 level-wise growth when the total number of splits is identical. Additionally, it reduces redundant computations
205 involving leaf nodes with low gain, thereby guaranteeing the model's computational efficiency without
206 compromising its accuracy.

207 **3.3.3 XGBoost**

208 XGBoost, a robust machine learning framework designed to enhance the scalability of tree-based boosting
209 algorithms, has emerged as a leading and widely adopted method in recent years across various research and
210 application domains (Chen et al.,2015; Zhang e et al.,2023). By leveraging gradient descent optimization to minimize
211 the predefined loss function, the algorithm constructs a predictive model through an ensemble of weak
212 classification trees within a boosting framework. The core idea lies in fitting well the residual of the previous
213 prediction by learning a new function in each iteration; on this basis, it further calculates the score corresponding
214 to each node in accordance with the sample characteristics.

215 In this study, three machine learning algorithms were employed to model the two scenarios (static factor only
216 and dynamic-static factors coupling) separately, resulting in a total of six models.

217 **3.4 Verification and comparison**

218 The ROC curve is a conventional method for validating the outcomes of LSA. The area under the ROC curve



219 (AUC, ranging from 0~1) is used to quantify the predictive accuracy of the assessment (Hodasová et al.,2021; Liang
220 et al.,2021; Liang et al.,2023).

221 The landslide susceptibility map (LSM) was reclassified into five levels as very low(0-0.2), Low(0.2-0.4),
222 moderate(0.4-0.6), high(0.6-0.8), very high (0.8-1). To verify the reliability of the susceptibility map, historical
223 landslide points and the size distribution of areas for each susceptibility level were applied (Liang et al.,2021).

224 The AUC, sensitivity, specificity, and LSM were applied to verify and compare the effectiveness of different
225 models in the study.

226 4. Results

227 4.1 Surface Deformation Results from SBAS - InSAR

228 Using the SBAS - InSAR method, we obtained the average annual surface deformation rate map of the study area
229 from January 2024 to December 2025. The results show that the surface deformation in the study area is spatially
230 heterogeneous, with deformation rates ranging from -8.1cm/year (subsidence) to 3 cm/year (uplift). Most areas
231 have stable surface conditions (deformation rate < 2cm/year, the green area in Fig.6), accounting for
232 approximately 75% of the total study area. The moderately unstable areas (deformation rate: -1 to -2 cm/year, the
233 yellow area in Fig.6) account for 20%, while the highly unstable areas (-2 to -8 cm/year, the red area in Fig.4)
234 account for 5%.

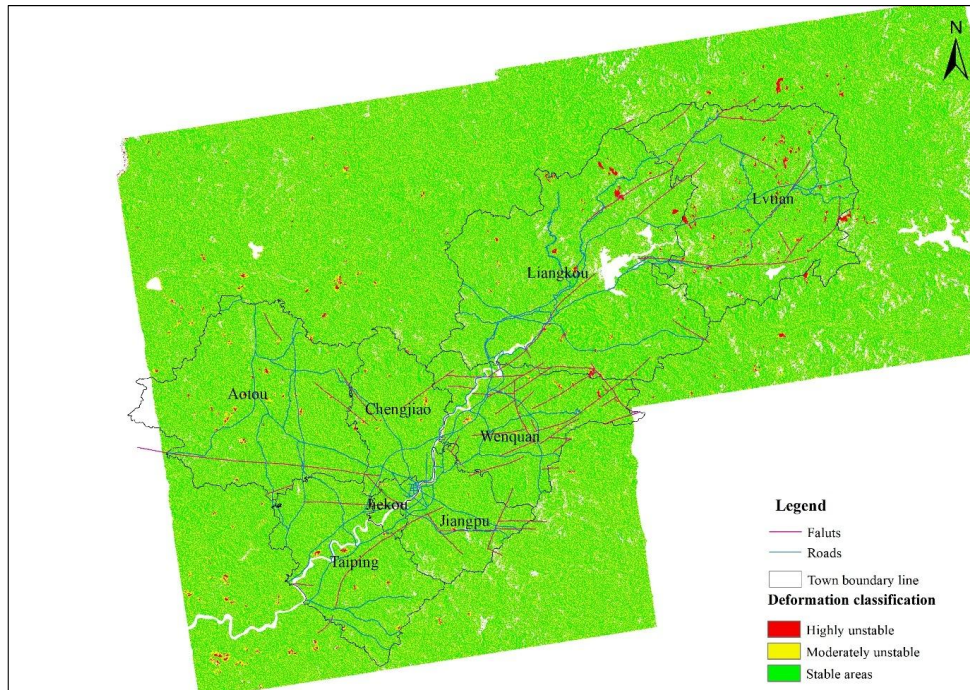
235 The highly unstable areas are mainly distributed in the following locations:

236 1. It mainly focuses on the areas of Lvian Town, the northern part of Liangkou Town, and the northern part
237 of Wenquan Town.

238 2. Along the roads, where road construction has disturbed the slope stability, leading to obvious subsidence.

239 3. Near the fault zones in the eastern part of the study area, where the geological structure is unstable, leading
240 to continuous subsidence.

241 These highly unstable areas are consistent with the field investigation (Fig.7), indicating that the InSAR-
242 derived surface deformation data can effectively identify potential shallow landslide risk areas. Besides, the
243 majority of these potential geohazard risks are closely associated with human engineering activities—including
244 road construction, residential development, and industrial facility construction—and such hazards typically
245 exhibit a small spatial scale.



246

247 **Figure 6: Surface Deformation Rate Map of the Study Area.**



248

249

250 **Figure 7: Field investigation:(a) slope collapse on 2025.6.17;(b) slope collapse on 2025.6.18 ;(c) landslide on 2025.8.20;(d)**
251 **landslide on 2025.8.20.**



252 **4.2 Model's Performance**

253 The performance of the models was evaluated using confusion matrices, with the detailed results presented in
 254 Table 4~5 and Fig.8. For the training dataset, the **SAR+XGBoost** model achieved the highest sensitivity (0.908),
 255 specificity (0.903), accuracy (0.906), and AUC (0.955), followed by the SAR+LGBM model with the values of
 256 0.887, 0.879, 0.883, and 0.943. The RF model performed the worst with the values of 0.834, 0.820, 0.827, and
 257 0.897. While the performance of the LGBM model is average, with the values of 0.842, 0.843, 0.843, and 0.915.
 258 The AUC values of all models were all close to 0.9. Among them, the SAR + XGBoost model achieved 0.955,
 259 which indicates that the model has achieved an excellent effect in predicting landslide susceptibility. After the
 260 integration based on SAR technology, the overall prediction accuracy of the model has improved to a certain
 261 extent.

262 In terms of validation, the SAR+XGBoost model also outperformed the others, reaching 0.865, 0.868, 0.867,
 263 and 0.933, whereas the XGBoost obtained 0.808, 0.799, 0.803, and 0.882. Except for the SAR + XGBoost model,
 264 the AUC values of all other models are all below 0.9. The RF model also performed the worst with the values of
 265 0.751, 0.736, 0.743, and 0.791. While the performance of the LGBM model is average, with the values of 0.781,
 266 0.77, 0.776, and 0.841. Compared with the training set, the performance of all models has declined to varying
 267 degrees, and the AUC value of the RF model has dropped below 0.8 but improved to 0.821 after merging SAR
 268 technology. After the integration based on SAR technology, the overall prediction accuracy of the model has
 269 improved to a certain extent.

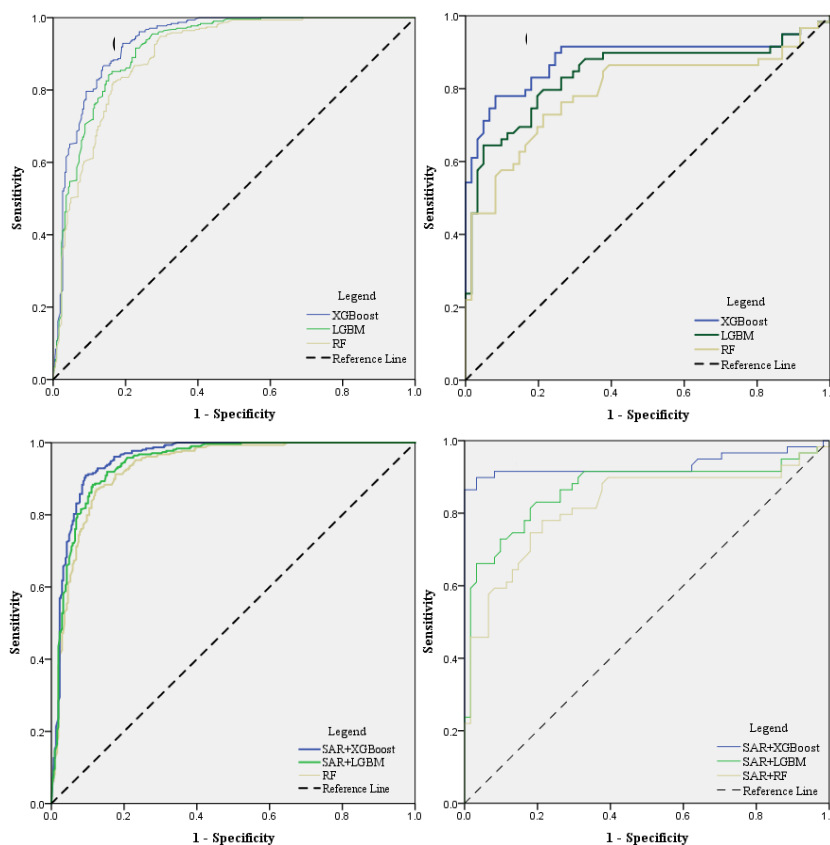
270 Compared to their performance on the training dataset, all models demonstrated varying degrees of
 271 performance degradation on the validation dataset. Specifically, the AUC value of the RF model dropped below
 272 0.8; however, after integrating SAR technology, its AUC value improved to 0.821.

273 **Table 4. Models' performance in training**

Indexs	XGBoost	LGBM	RF	SAR+XGBoost	SAR+LGBM	SAR+RF
TP	0.869	0.843	0.817	0.903	0.878	0.864
TN	0.849	0.842	0.837	0.909	0.889	0.878
FP	0.131	0.157	0.183	0.097	0.122	0.136
FN	0.151	0.158	0.163	0.091	0.111	0.122
Sensitivity	0.852	0.842	0.834	0.908	0.887	0.876
Specificity	0.866	0.843	0.820	0.903	0.879	0.866
Accuracy	0.859	0.843	0.827	0.906	0.883	0.871
AUC	0.934	0.915	0.897	0.955	0.943	0.932

274 **Table 5. Models' performance in validation**

Indexs	XGBoost	LGBM	RF	SAR+XGBoost	SAR+LGBM	SAR+RF
TP	0.797	0.766	0.727	0.869	0.794	0.766
TN	0.810	0.786	0.759	0.864	0.807	0.786
FP	0.203	0.234	0.273	0.131	0.206	0.234
FN	0.190	0.214	0.241	0.136	0.193	0.214
Sensitivity	0.808	0.781	0.751	0.865	0.804	0.781
Specificity	0.799	0.770	0.736	0.868	0.796	0.770
Accuracy	0.803	0.776	0.743	0.867	0.800	0.776
ROC	0.882	0.841	0.791	0.933	0.866	0.821



275

276

277 **Figure 8: Analysis of ROC curve: (a) traditional machine learning models in the training dataset; (b) traditional**
278 **machine learning models in the validation dataset; (c) SAR merged models in the training dataset; (d) SAR merged**
279 **models in the validation dataset.**

280 4.3 Landslide Susceptibility Map

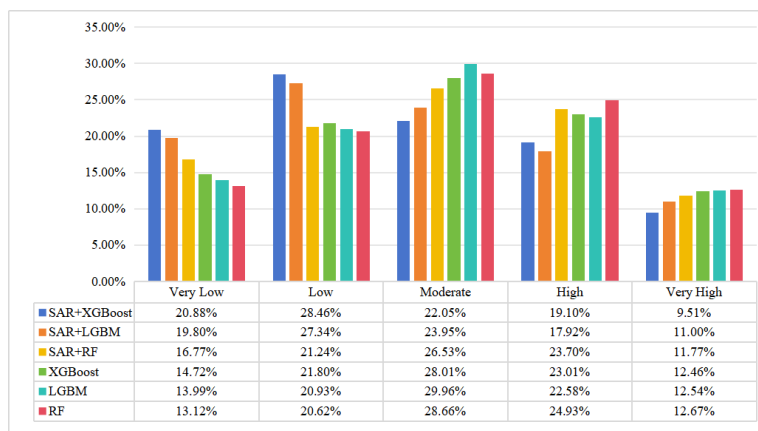
281 After model training and validation, the landslide susceptibility index (LSI) of each **slope unit** was calculated, and
282 the landslide susceptibility map of the whole study area was generated. The statistical results of the susceptibility
283 levels are shown in Fig.9. A logically sound landslide susceptibility map must adhere to two fundamental criteria:
284 (1) the density of landslide samples should exhibit a positive correlation with increasing susceptibility classes,
285 with the majority of landslide samples concentrated in the highest susceptibility zone; (2) the map should display
286 spatial continuity and smoothness, with the **very high-susceptibility class occupying a relatively small proportion**
287 **of the total study area** (Ado et al.,2022).

288 The traditional machine learning models consistently predict a relatively high proportion of moderately
289 susceptible areas, with all models yielding values exceeding 28%. In particular, the LGBM model achieved a rate
290 of approximately 30%. For the SAR+XGBoost model, the distribution of predicted susceptibility classes was:
291 20.88% (very low), 28.46% (low), 22.05% (moderate), 19.10% (high), and 9.51% (very high). While the XGBoost
292 model occupied 14.72% (very low), 21.80% (low), 28.01% (moderate), 23.01% (high), and 12.46% (very high),
293 respectively. The RF model occupied 13.12% (very low), 20.62% (low), 28.66% (moderate), 25.83% (high), and



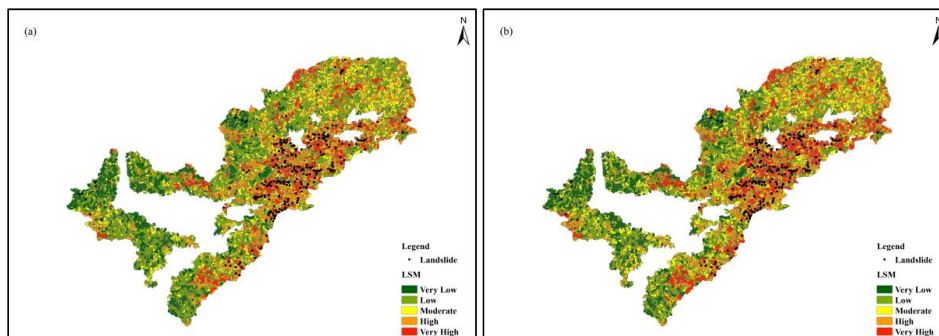
294 11.77(very high), respectively. While the SAR+RF model occupied 16.77% (very low), 21.24% (low), 26.53%
 295 (moderate), 23.70%(high), and 11.77% (very high), respectively.

296 **As illustrated in Fig.10, the landslide samples (represented by dark spots) are predominantly distributed in**
 297 **the red-colored regions (corresponding to high/very high susceptibility), while a few samples are scattered**
 298 **improperly in the green area (corresponding to low/very low susceptibility). Furthermore, the high and very high**
 299 **susceptibility zones are spatially concentrated in the eastern region, which is consistent with the actual spatial**
 300 **distribution of historical landslides.** This subregion is characterized by steep slopes (slope>30°) with substantial
 301 elevation differences (MED>300), close proximity to the adjacent road network (DTR<500), and a nearby
 302 distribution of geological faults. The low and very low susceptibility zones are spatially concentrated in the
 303 western region. This region is characterized by relatively gentle terrain with minimal elevation variations and a
 304 sparse distribution of faults. Collectively, these observations demonstrate that the susceptibility maps generated
 305 by the 6 models are generally logically consistent and reliable, but there are certain differences. Furthermore, as
 306 illustrated in Fig. 11, the high-deformation zones identified via SAR remote sensing do not fully overlap with the
 307 high-susceptibility areas delineated by the model; correspondingly, the stable zones detected by SAR technology
 308 also exhibit incomplete spatial overlap with the low-susceptibility areas. The highly unstable area is located in the
 309 plain area of the residential zone.



310

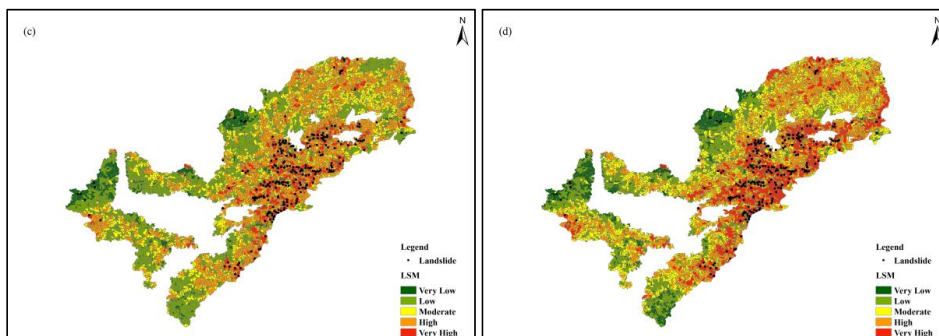
311 **Figure 9: Percentages of areas in LSM.**



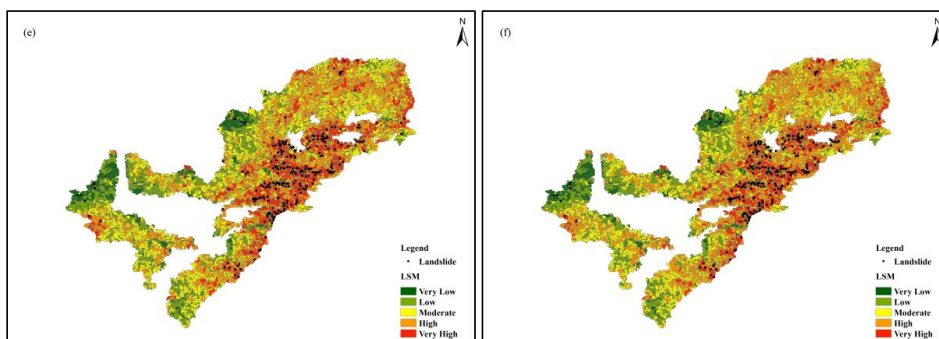
312



313



314

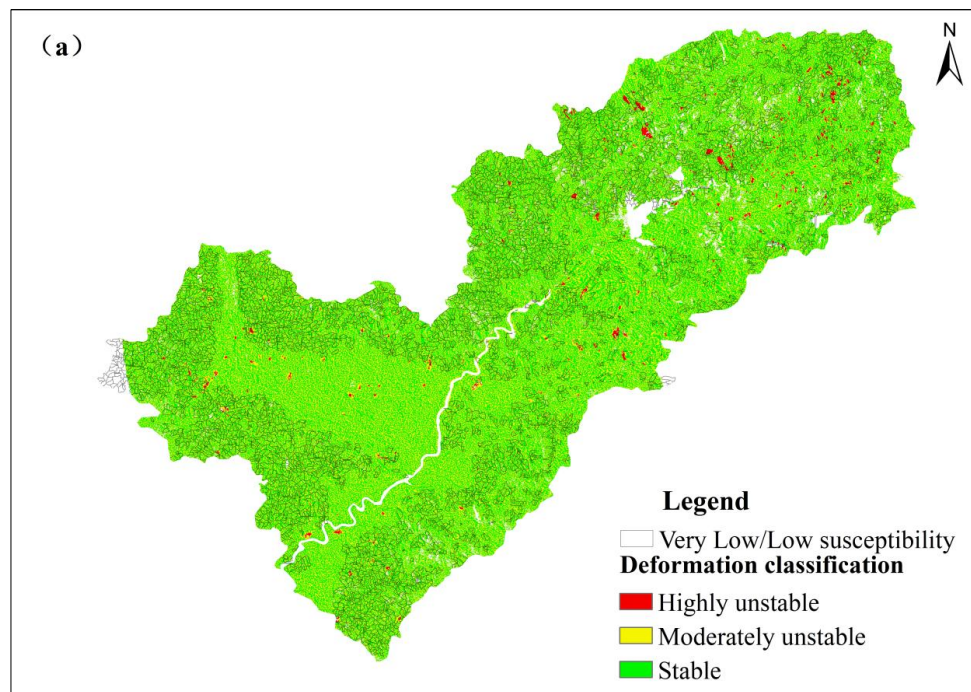


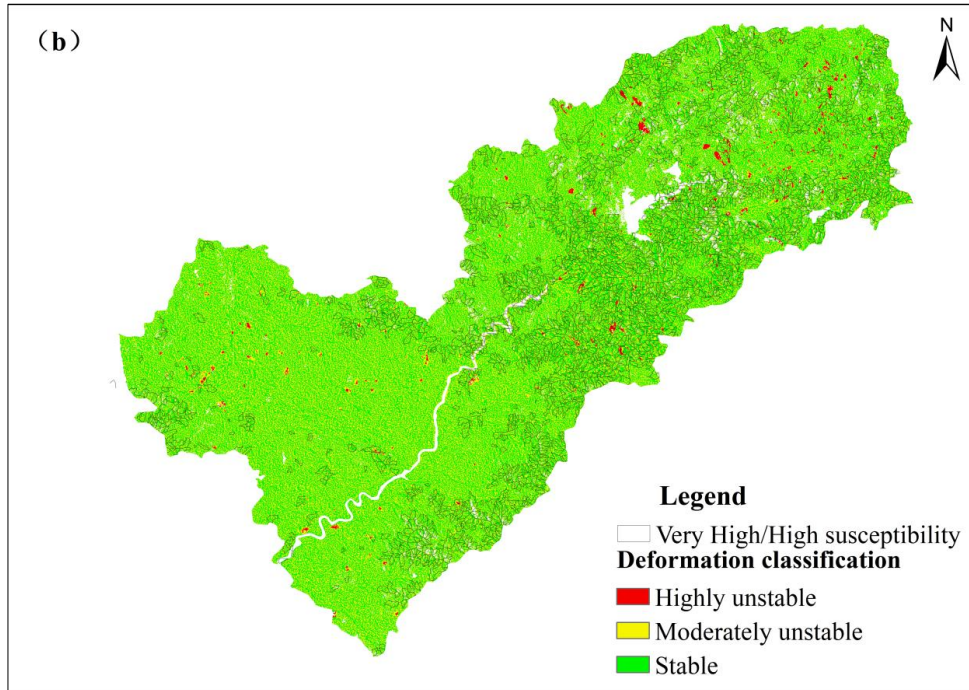
315

Figure 10: LSM produced by: (a) SAR+XGBoost; (b) SAR+LGBM; (c) SAR+RF; (d) XGBoost; (e) LGBM; (f) RF.

316

317





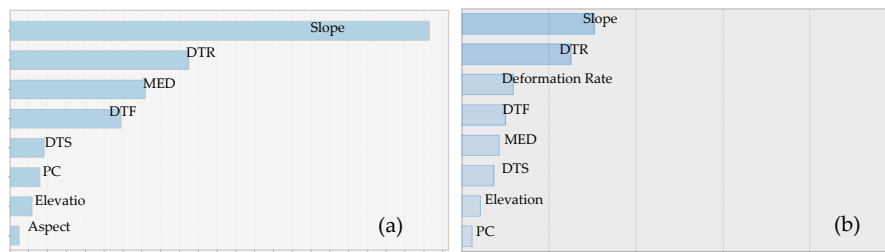
318

319 **Figure 11: LSM produced by SAR+XGBoost: (a) Low/very low susceptibility; (b) High/very high susceptibility.**

320 **4.4 Analysis of the Importance of Conditioning Factors**

321 To quantify the critical contribution of dynamic factors to landslide susceptibility prediction, the Gini index—an
 322 indicator that measures the correlation between predictive variables and model outputs—was employed in this
 323 study, and the results were normalized (Dou et al.,2023; Ye et al.,2022). In the static factor analysis, the key
 324 contributing variables were identified as the **slope, DTR, MED, and DTF**, with these factors exhibiting relatively
 325 high weight coefficients of 0.45, 0.19, 0.14, and 0.12, respectively (Fig.12; Table 6). By contrast, **DTS, PC,**
 326 elevation, and aspect yielded lower weight coefficients, indicating that these factors exert a relatively minor
 327 influence on landslide occurrence.

328 After coupling with SAR technology, the key contributing variables were identified as the slope, DTR,
 329 **Deformation rate (DR),** and DTF, with these factors exhibiting relatively high weight coefficients of 0.3, 0.25,
 330 0.12, and 0.1, respectively (Fig.12; Table 7). By contrast, MED, DTS, elevation, and PC yielded lower weight
 331 coefficients, indicating that these factors exert a relatively minor influence on landslide occurrence.



332



333 **Figure 12: Factor Importance Analysis: (a) Static factor analysis; (b) Dynamic and static factor coupling analysis.**

334 **Table 6 Static Factor Importance Analysis**

Factors	Slope	DTR	MED	DTF	DTS	PC	Elevation	Aspect
Gini	0.45	0.19	0.14	0.12	0.04	0.03	0.02	0.01

335 **Table 7 Dynamic and static factor Importance Analysis**

Factors	Slope	DTR	DR	DTF	MED	DTS	Elevation	PC
Gini	0.3	0.25	0.12	0.1	0.09	0.07	0.04	0.02

336 5. Discussion

337 5.1 Models' comparison

338 Different algorithms exhibit distinct focuses and design priorities, and their performance typically varies
 339 depending on the specific research context, characteristics of the study area, and quality of input samples (Hussin
 340 et al.,2016; Lai et al.,2019; Yilmaz et al.,2010). With the advancement of related technologies and the maturation of
 341 theoretical frameworks, a variety of advanced algorithms—such as convolutional neural networks (CNNs)—have
 342 been developed and increasingly applied to landslide susceptibility mapping (LSM) (Wang et al.,2019; Zhao et
 343 al.,2024). These algorithmic advancements aim to enhance the prediction accuracy and reliability of LSM results.
 344 The three algorithms adopted in this study are all tree-based ensemble algorithms. Regarding accuracy, the
 345 performance of the three algorithms was similar throughout the training and validation stages. Overall, the
 346 integration of SAR technology contributed to a measurable improvement in the general prediction accuracy of the
 347 models. Tree-based ensemble algorithms are capable of efficiently learning discriminative features from SAR-
 348 derived ground deformation interpretation data, illustrating the utility of incorporating this remote sensing
 349 technology to enhance the generalization capability of landslide susceptibility assessment models.

350 For LSM, the six models yield generally consistent predictions of the overall susceptibility spatial trends.
 351 However, traditional machine learning methods tend to delineate a larger proportion of the study area as
 352 moderately susceptible zones, which increases the uncertainty of the prediction results. Models integrated with
 353 SAR technology exhibit a reduced proportion of high/very high landslide susceptibility areas, accompanied by a
 354 corresponding increase in the proportion of low and very low susceptibility areas.

355 5.2 Contribution of InSAR-Derived Deformation Data to Susceptibility Assessment


356 In this study, ground deformation is treated as a dynamic influencing factor and incorporated into the
 357 comprehensive evaluation framework for landslide susceptibility. Compared with models that solely incorporate
 358 static factors, the algorithms integrated with SAR technology not only achieve improved prediction accuracy but
 359 also yield more rational spatial distribution patterns across different landslide susceptibility classes (Hua et al.,2021;
 360 Lamichhane et al.,2025). Through the analysis of the importance of the conditioning factors, deformation rate was
 361 essential (Devara et al.,2021).

362 Most prior studies rely on detailed landslide inventories, operating under the implicit assumption that the
 363 spatiotemporal patterns of future landslides will be consistent with historical and contemporary observations
 364 (Steger et al.,2017; Huang et al.,2024; Zhi et al.,2023). However, intensive human activities and sustained rainfall



365 events frequently induce novel landslide patterns, thereby challenging the validity of this fundamental premise.
366 The deployment of landslide monitoring equipment is not only associated with substantial economic costs but
367 also faces challenges in providing effective early warning for sudden-onset geohazards. Notably, in southern
368 China, slope cutting activities associated with rural residential construction and road development are widespread,
369 and geological hazards are characterized by high frequency, extensive spatial distribution, and predominantly
370 small scales, rendering comprehensive coverage by monitoring systems technically infeasible and economically
371 impractical (Casagli et al.,2023).

372 Static landslide susceptibility assessment models tend to overestimate the stability of slopes that have been
373 engineered for stabilization, while underestimating the instability of slopes disturbed by anthropogenic activities
374 (Singh et al.,2010). Furthermore, the calibration of static landslide susceptibility assessment models must be
375 updated in tandem with the refinement and expansion of landslide inventories. However, the refinement and
376 expansion of landslide inventories are both financially costly and logistically infrequent. Consequently, the
377 absence of timely updated landslide inventories readily leads to the underestimation of emerging landslide patterns
378 triggered by urban expansion or shifts in meteorological event regimes.

379 Compared with traditional assessment methods that ignore dynamic deformation information, the integration
380 of InSAR data provides two key improvements. 1. **Early Identification of Potential Risks:** Traditional methods
381 mainly rely on static factors (e.g., lithology, slope) to identify high-risk areas, but these factors cannot reflect the
382 dynamic changes in slope stability. **The InSAR-derived deformation data can capture subtle subsidence (e.g., -5**
**to -10 mm/year) that is imperceptible to field surveys, enabling early identification of slopes in the "pre-landslide"**
384 stage; 2. **Improvement of Assessment Accuracy:** The validation results show that the overall accuracy of the
385 susceptibility map is **15 - 20%** higher than that of traditional static assessment methods. This significant
386 improvement confirms that InSAR data effectively fills the gap of dynamic information in traditional assessments
387 (Chen et al.,2023).

388 **However, the high- and low-deformation zones identified via SAR remote sensing do not precisely**
389 correspond to the high- and low-susceptibility areas predicted by the landslide susceptibility model. SBAS -
390 InSAR also has limitations. For example, in areas with extremely high vegetation coverage (e.g., dense primeval
391 forests), the coherence of SAR images is still low, leading to missing deformation data. Conversely, geological
392 hazard occurrences exhibit inherent spatiotemporal randomness, while the current SAR remote sensing systems
393 are constrained by finite spatiotemporal resolution. Therefore, the evaluation of landslide susceptibility should
394 incorporate both static and dynamic factors.

395 6. Conclusions

396 A high-quality landslide susceptibility map provides clear guidance for local disaster prevention and mitigation
397 work. This study systematically explored the application of SBAS-InSAR in small-scale landslide susceptibility
398 dynamic assessment, using tree-based ensemble algorithms and the Conghua district as the study area. The main
399 conclusions are as follows:

400 **1. The SBAS-InSAR technique enables the effective capture of spatiotemporal dynamic surface deformation**
401 information, which is of considerable significance for the identification and monitoring of small-scale landslide
402 hazards in the southern China region.

403 2. Following the integration of deformation-derived dynamic factors, LSA based on the tree-based algorithm



404 has achieved significant improvements in prediction accuracy and reliability—demonstrating the applicability of
405 fusing SAR-derived deformation data with machine learning algorithms for dynamic landslide risk evaluation.

406 However, there are some limitations of the Study:

407 1. Diverse SAR technologies (e.g., SBAS-InSAR, PS-InSAR, or TomoSAR) can be leveraged to extract
408 higher-dimensional deformation features, which can be incorporated into LSA frameworks.

409 2. The integrated technology requires rigorous validation through multiple heterogeneous algorithms and
410 cross-validation across distinct geomorphological study areas with varying geological.

411 Data availability

412 The data used to support the findings of this study are included within the article.

413 Author Contributions

414 Z.L., writing—original draft, methodology, and software; W.P., W.L. and H.H., review and validation; J.H. and
415 K.L., reviewing and editing. K.J. and G.L. are responsible for the review and validation. All authors have read
416 and agreed to the published version of the manuscript.

417 Competing interests

418 The authors declare that they have no conflict of interest.

419 Financial support

420 The completion of this work was supported by Collaborative Innovation Center for Natural Resources Planning
421 and Marine Technology of Guangzhou (No. 2023B04J0301, No. 2025B04J0031), the Science and Technology
422 Foundation of Guangzhou Urban Planning & Design Survey Research Institute (Grant No. RDI2240204004). Last
423 but not least, many thanks for the support of Academic Specialty Group for Urban Sensing in Chinese Society of
424 Urban Planning.

425 References

- 426 1. Ado M, Amitab K, Maji A K, et al. Landslide susceptibility mapping using machine learning: A literature survey[J].
427 Remote Sensing, 2022, 14(13): 3029.
- 428 2. Castelli M, Filipello A, Fasciano C, et al. Geohazard Plugin: A QGIS Plugin for the Preliminary Analysis of
429 Landslides at Medium–Small Scale[J]. Land, 2025, 14(2): 290.
- 430 3. Sun L, Zhao Y, Sun Y, et al. Study on the reactivation mechanism and weak substrate effect of bottom-saturated
431 reservoir old landslides—based on a small-scale model[J]. Bulletin of Engineering Geology and the Environment, 2025,
432 84(9): 420.
- 433 4. Zhang K, Song H, Sun S, et al. Distribution and genesis of submarine landslides in the northeastern South China
434 Sea[J]. Geological Journal, 2021, 56(3): 1187-1201.
- 435 5. Wang D, Hao M, Chen S, et al. Assessment of landslide susceptibility and risk factors in China[J]. Natural hazards,
436 2021, 108(3): 3045-3059.
- 437 6. Cui Y, Qian Z, Xu W, et al. A Small-Scale Landslide in 2023, Leshan, China: Basic Characteristics, Kinematic
438 Process and Cause Analysis[J]. Remote Sensing, 2024, 16(17): 3324.
- 439 7. Kovács I P, Czirány S, Dobre B, et al. A field survey–based method to characterise landslide development: a case



- 440 study at the high bluff of the Danube, south-central Hungary[J]. *Landslides*, 2019, 16(8): 1567-1581.
- 441 8. Huang S, Chen L. Landslide susceptibility mapping using an integration of different statistical models for the 2015
442 Nepal earthquake in Tibet[J]. *Geomatics, Natural Hazards and Risk*, 2024, 15(1): 2396908.
- 443 9. Nwazelibé V E, Egbueri J C. Geospatial assessment of landslide-prone areas in the southern part of Anambra State,
444 Nigeria using classical statistical models[J]. *Environmental Earth Sciences*, 2024, 83(7): 220.
- 445 10. Huang F, Xiong H, Jiang S H, et al. Modelling landslide susceptibility prediction: A review and construction of semi-
446 supervised imbalanced theory[J]. *Earth-Science Reviews*, 2024, 250: 104700.
- 447 11. Hong H, Wang D, Zhu A X, et al. Landslide susceptibility mapping based on the reliability of landslide and non-
448 landslide sample[J]. *Expert Systems with Applications*, 2024, 243: 122933.
- 449 12. Zhou C, Gan L, Cao Y, et al. Landslide susceptibility assessment of the Wanzhou district: Merging landslide
450 susceptibility modelling (LSM) with InSAR-derived ground deformation map[J]. *International Journal of Applied*
451 *Earth Observation and Geoinformation*, 2025, 136: 104365.
- 452 13. Wang C, Yang Y, Zhang X, et al. Ultra-High Sensitivity Real-Time Monitoring of Landslide Surface Deformation via
453 Triboelectric Nanogenerator[J]. *Advanced Materials*, 2024, 36(50): 2410471.
- 454 14. Ebrahim K M P, Gomaa S M M H, Zayed T, et al. Recent phenomenal and investigational subsurface landslide
455 monitoring techniques: a mixed review[J]. *Remote Sensing*, 2024, 16(2): 385.
- 456 15. Nie W, Tian C, Song D, et al. Disaster process and multisource information monitoring and warning method for
457 rainfall-triggered landslide: a case study in the southeastern coastal area of China[J]. *Natural Hazards*, 2025, 121(3):
458 2535-2564.
- 459 16. Wang L, Zhang K, Chen Y, et al. Progressive deformation mechanism of colluvial landslides induced by rainfall:
460 insights from long-term field monitoring and numerical study[J]. *Landslides*, 2024, 21(12): 3069-3086.
- 461 17. Hussain S, Pan B, Hussain W, et al. Integrated PSInSAR and SBAS-InSAR analysis for landslide detection and
462 monitoring[J]. *Physics and Chemistry of the Earth, Parts A/B/C*, 2025: 103956.
- 463 18. Zhang J, Zuo X, Li Y, et al. Detection and assessment of potential landslides in the Xiaojiang River Basin using
464 SBAS-InSAR[J]. *Scientific Reports*, 2025, 15(1): 16082.
- 465 19. Guo J, Xi W, Yang Z, et al. Landslide hazard susceptibility evaluation based on SBAS-InSAR technology and SSA-
466 BP neural network algorithm: A case study of Baihetan Reservoir Area[J]. *Journal of Mountain Science*, 2024, 21(3):
467 952-972.
- 468 20. Berardino, P., Fornaro, G., Lanari, R., & Sansosti, E. (2002). A new algorithm for surface
469 deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions*
470 *on Geoscience and Remote Sensing*, 40(11), 2375 - 2383.
- 471 21. Lie X U, Nengpan J U, Mingdong D, et al. Comparative analysis of landslide susceptibility assessment accuracy in
472 different evaluation units[J]. *Journal of Engineering Geology*, 2024, 32(5): 1640-1653.
- 473 22. Liu S, Du J, Yin K, et al. Regional early warning model for rainfall induced landslide based on slope unit in
474 Chongqing, China[J]. *Engineering Geology*, 2024, 333: 107464.
- 475 23. Li M, Wang H, Chen J, et al. Assessing landslide susceptibility based on the random forest model and multi-source
476 heterogeneous data[J]. *Ecological Indicators*, 2024, 158: 111600.
- 477 24. Breiman L. Random forests[J]. *Machine learning*, 2001, 45(1): 5-32.
- 478 25. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[J]. *Advances in neural*
479 *information processing systems*, 2017, 30.
- 480 26. Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting[J]. *R package version 0.4-2*, 2015, 1(4): 1-4.
- 481 27. Zhang J, Ma X, Zhang J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-
482 XGBoost model[J]. *Journal of environmental management*, 2023, 332: 117357.
- 483 28. Hodasová K, Bednarik M. Effect of using various weighting methods in a process of landslide susceptibility
484 assessment[J]. *Natural Hazards*, 2021, 105(1): 481-499.
- 485 29. Liang Z, Wang C, Duan Z, et al. A hybrid model consisting of supervised and unsupervised learning for landslide
486 susceptibility mapping[J]. *Remote Sensing*, 2021, 13(8): 1464.
- 487 30. Liang Z, Peng W, Liu W, et al. Exploration and comparison of the effect of conventional and advanced modeling
488 algorithms on landslide susceptibility prediction: A case study from Yadong Country, Tibet[J]. *Applied Sciences*,
489 2023, 13(12): 7276.
- 490 31. Liang Z, Wang C, Khan K U J. Application and comparison of different ensemble learning machines combining with a
491 novel sampling strategy for shallow landslide susceptibility mapping[J]. *Stochastic Environmental Research and Risk*
492 *Assessment*, 2021, 35(6): 1243-1256.
- 493 32. Ado M, Amitab K, Maji A K, et al. Landslide susceptibility mapping using machine learning: A literature survey[J].
494 *Remote Sensing*, 2022, 14(13): 3029.
- 495 33. Dou H, He J, Huang S, et al. Influences of non-landslide sample selection strategies on landslide susceptibility



- 496 mapping by machine learning[J]. *Geomatics, Natural Hazards and Risk*, 2023, 14(1): 2285719.
497 34. Ye C, Wei R, Ge Y, et al. GIS-based spatial prediction of landslide using road factors and random forest for Sichuan-
498 Tibet Highway[J]. *Journal of Mountain Science*, 2022, 19(2): 461-476.
499 35. Hussin H Y, Zumpano V, Reichenbach P, et al. Different landslide sampling strategies in a grid-based bi-variate
500 statistical susceptibility model[J]. *Geomorphology*, 2016, 253: 508-523.
501 36. Lai J S, Chiang S H, Tsai F. Exploring influence of sampling strategies on event-based landslide susceptibility
502 modeling[J]. *ISPRS International Journal of Geo-Information*, 2019, 8(9): 397.
503 37. Yilmaz I. The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and
504 artificial neural networks[J]. *Environmental Earth Sciences*, 2010, 60(3): 505-519.
505 38. Wang Y, Fang Z, Hong H. Comparison of convolutional neural networks for landslide susceptibility mapping in
506 Yanshan County, China[J]. *Science of the total environment*, 2019, 666: 975-993.
507 39. Zhao Z, Chen T, Dou J, et al. Landslide susceptibility mapping considering landslide local-global features based on
508 CNN and transformer[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024,
509 17: 7475-7489.
510 40. Hua Y, Wang X, Li Y, et al. Dynamic development of landslide susceptibility based on slope unit and deep neural
511 networks[J]. *Landslides*, 2021, 18(1): 281-302.
512 41. Lamichhane S, Kansakar A R, Devkota N, et al. Integrating dynamic factors for predicting future landslide
513 susceptibility[J]. *Environmental Earth Sciences*, 2025, 84(3): 89.
514 42. Devara M, Tiwari A, Dwivedi R. Landslide susceptibility mapping using MT-InSAR and AHP enabled GIS-based
515 multi-criteria decision analysis[J]. *Geomatics, Natural Hazards and Risk*, 2021, 12(1): 675-693.
516 43. Steger S, Brenning A, Bell R, et al. The influence of systematically incomplete shallow landslide inventories on
517 statistical susceptibility models and suggestions for improvements[J]. *Landslides*, 2017, 14(5): 1767-1781.
518 44. Huang F, Mao D, Jiang S H, et al. Uncertainties in landslide susceptibility prediction modeling: A review on the
519 incompleteness of landslide inventory and its influence rules[J]. *Geoscience Frontiers*, 2024, 15(6): 101886.
520 45. Zhiyong F, Changdong L, Wenmin Y. Landslide susceptibility assessment through TrAdaBoost transfer learning
521 models using two landslide inventories[J]. *Catena*, 2023, 222: 106799.
522 46. Casagli N, Intrieri E, Tofani V, et al. Landslide detection, monitoring and prediction with remote-sensing
523 techniques[J]. *Nature Reviews Earth & Environment*, 2023, 4(1): 51-64.
524 47. Singh T N, Verma A K, Sarkar K. Static and dynamic analysis of a landslide[J]. *Geomatics, Natural Hazards and Risk*,
525 2010, 1(4): 323-338.
526 48. Chen L, Ma P, Yu C, et al. Landslide susceptibility assessment in multiple urban slope settings with a landslide
527 inventory augmented by InSAR techniques[J]. *Engineering Geology*, 2023, 327: 107342.

528