RESEARCH PAPER

LANDSLIDE SUSCEPTIBILITY PREDICTION ON MOUNT MARAPI USING INTERFEROMETRIC SYNTHETIC APERTURE RADAR (INSAR) INTEGRATED WITH MULTI-MODEL MACHINE LEARNING APPROACHES

Muhammad Hanif[1*], Sarun Apichontrakul[1], Rini Suryani[2], Ravidho Ramadhan[3]
[1]Geoinformatics, College of Computing, Khon Kaen University, Mittraphap Road, Muang District, Khon Kaen City, 40002, Thailand
[2]Department of Civil Engineering, Khon Kaen University, Mittraphap Road, Muang District, Khon Kaen City, 40002, Thailand
[3]Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji City, Kyoto Prefecture, 611-0011, Japan
*Corresponding author: muhammadhanif.m@cassia.kku.ac.th
Received: April 23rd 2025 / Accepted: March 8th 2026 / Published: June 30th 2026
https://doi.org/10.24057/2071-9388-2026-4016

ABSTRACT. Landslides on Mount Marapi, West Sumatra, Indonesia, pose significant threats to human life, infrastructure, and the surrounding environment. This s.tudy examines patterns of surface deformation and predicts landslide susceptibility on Mount Marapi by integrating Interferometric Synthetic Aperture Radar (InSAR) interferograms with a multi-model machine-learning framework. Landslide susceptibility prediction was conducted using K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB) algorithms. Integration of InSAR-derived deformation data with geospatial variables and machine learning facilitates a more comprehensive understanding of deformation dynamics and landslide susceptibility. The findings reveal a dominant trend of subsidence accompanied by persistent fluctuations along the mountain slopes, largely driven by recurrent landslides across multiple locations, thereby reshaping the surface morphology. InSAR data provide a spatially continuous representation of deformation within landslide-prone zones. Among the evaluated algorithms, RF demonstrated the highest predictive performance, yielding an accuracy of 94.33% and a precision of 91.30%. Key variables such as slope gradient, elevation, and rainfall emerged as the most influential determinants of landslide susceptibility. Additionally, seismic activity along the subduction zone and tectonic earthquakes act as further triggers, intensifying landslide occurrence in conjunction with extreme rainfall events. Although secondary factors such as curvature, soil type, and localized deformation exert relatively weaker influences, their interactions remain important in constructing a holistic landslide prediction framework. Overall, this study highlights the effectiveness of integrating InSAR with machine learning to advance multivariable approaches for forecasting and mitigating landslide hazards on Mount Marapi.

KEYWORDS: DInSAR, ground deformation, landslide, GeoAI, machine learning

CITATION: Hanif M., Apichontrakul S., Suryani R., Ramadhan R. (2026). Landslide Susceptibility Prediction On Mount Marapi Using Interferometric Synthetic Aperture Radar (Insar) Integrated With Multi-Model Machine Learning Approaches. Geography, Environment, Sustainability, 2 (19), 29-40
https://doi.org/10.24057/2071-9388-2026-4016

ACKNOWLEDGEMENTS: The authors wish to express gratitude to Khon Kaen University. We also thank the providers of all open-source data used in this study, as well as Mr. Rizki Atthoriq Hidayat and Mr. Triyatno for their support of this research.

Conflict of interests: The authors reported no potential conflict of interests.

INTRODUCTION

Natural disasters, particularly landslides in mountainous regions, represent severe threats to human life, infrastructure, and ecosystems. Indonesia, located along the Pacific Ring of Fire, is highly prone to geological hazards, with volcanic activity constituting a major source of risk. Following the major eruption of Mount Marapi in December 2023, the surrounding area of West Sumatra experienced a subsequent disaster in May 2024: a landslide triggered by the volcano's geomorphic and environmental conditions (Abu et al. 2021). Meteorological conditions, particularly intense rainfall, represent a primary triggering factor for landslide movements and cold lava floods, leading to severe damage, loss of life, and disruption of local livelihoods (Wulansari et al., 2024). The presence of unconsolidated volcanic deposits on steep, unstable slopes further amplifies susceptibility to mass movements during heavy precipitation events (Narny et al., 2025). In addition to hydrometeorological triggers, geophysical factors such as proximity to tectonic fault systems, seismic activity, and variations in peak ground acceleration following earthquakes play a critical role in destabilizing the landscape, particularly in volcanic terrains (Le Breton et al. 2021). Despite the recurrence of landslides in the Mount Marapi region, comprehensive mapping and predictive modeling efforts remain limited. Addressing this gap is vital for understanding the complex interplay of geomorphic and geophysical processes in one of Indonesia's most tectonically and volcanically active regions, thereby improving disaster preparedness and risk mitigation strategies.

Synthetic Aperture Radar (SAR)-based active remote sensing, which employs microwave signals, has proven to be a highly effective tool for detecting and monitoring ground deformation associated with landslides and other mass-movement processes (Mondini et al. 2019). Particularly noteworthy is the use of Sentinel-1 time-series data processed through the LiCSAR automated system, which applies the Small Baseline Subset (SBAS) InSAR approach to generate detailed temporal and spatial information on surface deformation (Lazecky, 2020). Leveraging Interferometric Synthetic Aperture Radar (InSAR) techniques allows for the detection of subtle ground displacements with high precision, thereby facilitating the identification of precursory deformation in landslide-prone areas. This capability is critical for the early recognition of hazardous zones, as regions exhibiting prior deformation are considerably more susceptible to subsequent slope failures. Such geospatial intelligence plays a pivotal role in informing proactive disaster mitigation strategies, particularly within volcanic environments (Hanif et al., 2024a).

Advancements in machine learning (ML) applied to geospatial data, commonly referred to as Geo-Artificial Intelligence (GeoAI) has introduced transformative opportunities for natural hazard prediction and risk reduction (Demertzis et al. 2021). GeoAI represents an interdisciplinary domain that integrates artificial intelligence (AI), geographic information systems (GIS), and geospatial science to intelligently analyze, model, and interpret complex spatial datasets. By employing AI methodologies such as machine learning, deep learning, and neural networks, GeoAI enables the efficient processing of large-scale geospatial data, facilitating the extraction of meaningful patterns and actionable insights (Lu et al. 2025).

The application of GeoAI in disaster studies, particularly landslide prediction, has demonstrated significant potential. Machine learning algorithms are well-suited for analyzing multidimensional spatial datasets, detecting nonlinear relationships, and producing robust forecasts of hazardous events. Within the context of landslide risk assessment, ML-based models substantially improve predictive accuracy, thereby supporting the delineation of high-risk zones. This capability is critical for informing proactive disaster mitigation strategies and enhancing resilience in vulnerable regions (Mao et al. 2024).

The present study addresses the research gap in post-eruption landslide analysis in volcanic environments by integrating remote sensing and machine learning approaches to systematically map and predict landslides in volcanic terrains. Beyond advancing scientific understanding, this research provides practical value for disaster risk governance by delivering evidence-based insights that strengthen early warning systems and guide policy interventions. Moreover, the study contributes to the broader geoscientific discourse on landslide dynamics in volcanic landscapes, offering a methodological framework applicable to other disaster-prone regions.

The objectives of this study are to analyze volcano surface deformation patterns at landslide event locations from 2020 to 2024 and to integrate deformation data with diverse spatial datasets and machine learning for predicting areas susceptible to landslides. By combining remote sensing technologies with advanced data-driven modeling approaches, this research contributes to disaster risk reduction in volcanic environments. The findings are expected to support the strengthening of local community resilience and the mitigation of socio-environmental impacts arising from future geological hazards in the Mount Marapi region.

MATERIALS AND METHODS

Sampling Location

Mount Marapi, located in West Sumatra, is recognized as the most active volcano on the island of Sumatra. Its name, which translates to "*Mountain of Fire*", aptly reflects its persistent volcanic activity. Rising to an elevation of 2,891 meters above sea level, Marapi is characterized by explosive eruptions that generate violent blasts of pyroclastic material, volcanic ash, and lava flows. From a geological perspective, the volcano is positioned along the Semangko Fault, a major tectonic structure within the Sunda subduction zone, extending from the Sianok Valley to Lake Singkarak (Fig. 1). Administratively, Mount Marapi lies within two regencies, Agam Regency and Tanah Datar Regency, under the jurisdiction of West Sumatra Province.

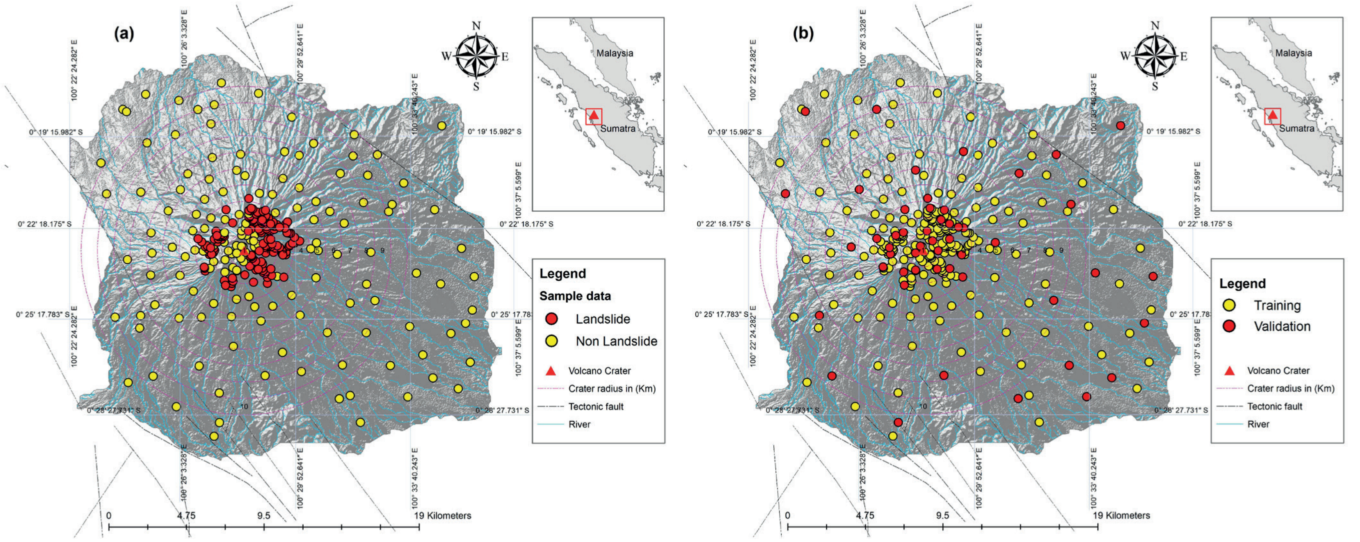

Fig. 1. (a) Distribution of landslide and non-landslide data, and (b) distribution of training and validation data in the study area at Marapi Volcano, West Sumatra, Indonesia

In this study, historical landslide occurrence points from 2020 to 2024 were compiled through on-screen digitization of high-resolution satellite imagery inventories using Google Earth. To establish a balanced dataset, an equal number of non-landslide points were randomly sampled from areas with no recorded landslide activity. Each landslide point was assigned a class value of 1, while non-landslide points were assigned a class value of 0. A total of 294 sample data points were compiled, with a 50/50 ratio of points labeled as landslides and non-landslides. These labeled points were subsequently used as reference data for training and validating the machine learning-based landslide prediction models, which were divided into training and validation sets in a 70/30 ratio.

Material

This study employs multiple open-access datasets, which were systematically integrated and processed into derivative products tailored to the analytical requirements. All data were obtained from official provider platforms, ensuring reliability and reproducibility. The datasets and their corresponding information are summarized in Table 1. To maintain consistency and facilitate spatial comparability, all raster datasets were resampled to a standardized spatial resolution of 50 meters.

Data preprocessing for the variable analysis was conducted using an integrated GIS and remote sensing framework. Land-cover classification was implemented through supervised classification techniques, while terrain characteristics were derived from Google Earth Engine. Earthquake density, gravity anomalies, and river density were quantified using kernel density estimation.

InSAR for Surface Deformation Monitoring

Sentinel-1 Synthetic Aperture Radar (SAR) data were processed using the LiCSAR (LiCS Automated Processing System) platform to generate Interferometric Synthetic Aperture Radar (InSAR) products. The data used are of descending type with temporal recording. LiCSAR processing includes multiple stages such as burst extraction, frame compilation, and image coregistration, incorporating spectral diversity correction and other essential adjustments (Lazecky, 2020). To address decorrelation in low-coherence areas, LiCSAR applies several techniques, including spatial and temporal filtering, multi-interferogram stacking, and advanced approaches such as Persistent Scatterer InSAR (PS-InSAR) and Small Baseline Subset (SBAS). In particular, SBAS time-series interferograms are generated through a small-baseline strategy and subsequently refined via iterative decomposition (Morishita, 2020).

For the detailed assessment of surface deformation at landslide locations, unwrapped phase time-series data were extracted from the InSAR products. Statistical analyses were then performed for each landslide sample point, including temporal trend evaluation, calculation of maximum and minimum displacement values, and deviation analysis. These deformation metrics provide critical insights into the spatial and temporal variability of displacement processes in landslide-prone areas, thereby enhancing the understanding of slope instability dynamics.

Machine Learning for Predicting Landslide Susceptibility

This study employs a multi-model machine learning framework based on supervised learning algorithms. The models operate as binary classifiers, requiring labeled training data in which a value of "1" denotes the presence of a landslide and "0" indicates its absence. Such labeled datasets allow the algorithms to recognize patterns and establish predictive relationships between input variables and landslide occurrences. To ensure a fair and objective comparison among the algorithms employed in this study, the hyperparameters of each machine learning model were systematically tuned to obtain their respective optimal configurations. The selected optimal parameters are then applied to the algorithms implemented in this study.

The selected machine learning algorithms are capable of producing both classification and regression outputs, depending on the predictive objective. While each algorithm applies distinct computational strategies, together they capture complex non-linear interactions among variables. In comparison with conventional statistical techniques, machine learning methods exhibit greater capacity to manage high-dimensional geospatial datasets and achieve higher predictive accuracy (Wu et al. 2024). The models implemented in this study include:

Table 1. Dataset of predictor variables for landslide susceptibility analysis

Data	Required information for predictor variables	Data Sources
Sentinel 1	(1) Surface deformation	https://comet.nerc.ac.uk/COMET-LiCS-portal/
Sentinel 2	(2) Land cover	https://www.copernicus.eu/en
Land systems	(3) Soil type	https://inaland.big.go.id/map
SRTM	(4) Elevation, (5) slope, (6) curvature, (7) river density	https://earthexplorer.usgs.gov/
Earthquake	(8) Earthquake density	https://earthquake.usgs.gov/earthquakes/search/
Seismic Hazard Map	(9) Peak Ground Acceleration (PGA)	https://www.globalquakemodel.org/product/global-seismic-hazard-map
Geology	(10) Geology, (11) distance to tectonic fault	https://geologi.esdm.go.id/geomap
Gravitation	(12) Gravitation	https://bgi.obs-mip.fr/
CHIRPS	(13) Precipitation	https://www.chc.ucsb.edu/data/chirps

• K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm classifies data points according to the majority class among their nearest neighbors, where the model determines neighborhood proximity by measuring the distance between a test sample and the corresponding training samples (Su et al. 2024). The KNN method functions as a non-parametric algorithm because it does not assume any predefined statistical distribution of the input data. This property allows the algorithm to adapt effectively to a wide range of geographic datasets with heterogeneous characteristics (Adnan et al. 2020). However, the performance of the KNN algorithm depends strongly on the selection of the parameter *k* and on the appropriate scaling of the input features. In addition, the computational cost of the algorithm increases substantially as the size of the dataset grows because the model must evaluate distances between the prediction sample and all training observations (Xia et al. 2017). Before the model performs predictions on the complete dataset, the study conducts a hyperparameter tuning procedure to identify the optimal configuration of the algorithm. The tuning process determines that the most suitable parameters for the model include the Euclidean distance metric, a value of *k* (n_neighbors) equal to 9, a *p* parameter value of 1, and a distance-based weighting scheme.

• Random Forest (RF)

The Random Forest (RF) algorithm employs a bagging strategy by constructing multiple Decision Tree classifiers on randomly selected subsets of the dataset and subsequently aggregating their predictions to produce a more robust final output. This ensemble learning approach reduces model variance, improves generalization performance, and enhances classification stability compared with the use of a single decision tree model (Li et al. 2024). In addition, the RF algorithm can effectively handle datasets containing a large number of variables and can provide valuable analytical insights through feature importance ranking, which is particularly beneficial when analyzing complex geospatial datasets (Chowdhury et al. 2024). However, the interpretation of aggregated predictions from the RF model is often less straightforward than the interpretation of a single decision tree structure (Wen et al. 2024). In this study, a hyperparameter tuning procedure was conducted to determine the optimal configuration of the RF model. The tuning process identified the following parameter settings as optimal: *n_estimators = 400, min_samples_split = 10, min_samples_leaf = 5, max_features = log2, max_depth* = None, and *bootstrap* = True. These optimized parameters were subsequently applied to train the model and perform predictions across the entire dataset used in this research.

• Support Vector Machine (SVM)

The Support Vector Machine (SVM) operates by constructing an optimal hyperplane that maximizes the margin separating data points belonging to different classes. The SVM model possesses a major advantage in its ability to address both linear and non-linear classification problems through the application of kernel functions. This capability makes the model particularly suitable for complex classification tasks involving spatial datasets (Huang et al. 2022). The SVM also demonstrates strong performance when handling high-dimensional datasets because the algorithm focuses on the most informative boundary samples, commonly referred to as support vectors (Xu et al. 2024). However, the effectiveness of the model remains highly sensitive to the selection of hyperparameters, and the computational cost can increase substantially when the algorithm is applied to large-scale datasets (Xia et al. 2017). In this study, the SVM model was trained prior to full implementation through a hyperparameter tuning procedure designed to identify the most suitable parameter configuration. The tuning results indicated that the optimal model configuration uses a sigmoid kernel, a gamma value of 0.1, and a regularization parameter *(C)* equal to 1. The study subsequently applied this optimized configuration to perform classification across the entire dataset used in this research.

• Gradient Boosting (GB)

The Gradient Boosting method represents an ensemble learning technique that constructs predictive models sequentially, where each subsequent model is trained to correct the residual errors generated by the preceding iteration. This learning strategy enables the model to gradually improve prediction performance by minimizing the remaining error at each stage of the training process (Kavzoglu and Teke, 2022). The Gradient Boosting approach is frequently applied to natural hazard prediction problems, including landslide susceptibility analysis (Zhang et al. 2022). In most implementations, the algorithm employs shallow Decision Tree structures as weak learners, allowing the ensemble model to incrementally enhance predictive accuracy through iterative error reduction (Sahin, 2022). Although the method is highly effective in capturing complex nonlinear relationships within datasets, the model requires careful hyperparameter tuning because excessive iterations or overly deep trees can increase the risk of overfitting (Sahin, 2022).

In this study, a hyperparameter tuning procedure was conducted to determine the most appropriate configuration of the Gradient Boosting model for the research dataset. The tuning results identified the following optimal parameter settings: *subsample = 0.7, n_estimators = 100, min_samples_split = 10, min_samples_leaf = 4, max_depth = 2, and learning_rate = 0.1*. The study subsequently applied this optimized configuration to train the model and perform predictive analysis on the dataset used in this research.

A multi-model machine learning framework was implemented by partitioning the dataset into training and testing subsets according to a predefined ratio (Abu et al. 2021), thereby reducing the risk of overfitting (see Fig. 1). Landslide prediction requires two principal datasets: (i) historical records of landslide events as a label dataset and (ii) predictor variable data. In this study, 13 predictor variables were selected (see Table 1).

Model performance was evaluated using multiple statistical metrics, including test accuracy, recall, F1-score, and Rate of Change (ROC) to ensure robust and reliable prediction outcomes. During the data exploration phase, correlation analysis was conducted to examine the interrelationships among variables, thereby identifying potential dependencies and assessing how variations in one factor may influence others. Furthermore, variable importance analysis was performed to quantify the contribution of each factor to landslide susceptibility, highlighting the most influential predictors for landslide occurrence in the Mount Marapi region. This step enhances the interpretability of the models and strengthens their predictive capability.

RESULTS

Trend surface ground deformation on the landslide

The results derived from InSAR time-series analysis, covering a five-year observation period, reveal distinct deformation patterns at landslide-prone locations based on the annual average deformation values (Table 2). Overall, deformation activity in the Mount Marapi region exhibited a declining trend between 2020 and 2024, although intermittent fluctuations with localized increases were still detected during certain intervals. The maximum annual deformation ranged from 53.70 mm in 2023 to 110.95 mm in 2022, with a five-year average of 79.30 mm. Conversely, the minimum deformation values ranged from –67.19 mm in 2021 to –157.58 mm in 2020, yielding an overall average of –74.42 mm across the study period. Meanwhile, the mean annual deformation values varied from 9.11 mm in 2024 to 20.64 mm in 2020, reflecting temporal variability in slope displacement intensity.

The deformation trend in the Mount Marapi region demonstrates a progressive decline in maximum deformation, decreasing from 110.95 mm in 2022 to 56.07 mm in 2024, thereby indicating a reduction in the intensity of surface displacement processes. In contrast, the minimum deformation values exhibit irregular fluctuations, with pronounced negative displacements of –67.19 mm in 2021 and –61.93 mm in 2023, reflecting localized subsidence phenomena. The standard deviation of the deformation data ranges from –36.13 mm to 14.48 mm, underscoring considerable temporal variability in ground movement. Notably, 2022 recorded the greatest variability, with a standard deviation of 14.48 mm, suggesting that deformation activity during this year was more spatially and temporally heterogeneous compared to other periods.

Spatial variations in the average deformation distribution over the five-year observation period are clearly discernible. Areas represented in red on the deformation map exhibit maximum uplift of up to 32.8 mm, whereas zones characterized by blue signatures indicate subsidence, with minimum values reaching –34.9 mm. Distinct spatial patterns emerge, particularly across the transitional zones between sloping and steep terrains extending toward the mountain summit. In contrast, the lowland plains at the mountain's base display comparatively moderate deformation values ranging from –17.9 mm to 15.9 mm, represented by a gradient of blue to orange hues (Fig. 2). This spatiotemporal deformation pattern constitutes a critical parameter for landslide susceptibility assessment in the Mount Marapi region. Furthermore, annual deformation data were systematically extracted for all recorded landslide events, as presented in Table 3.

Table 2. Annual deformation based on InSAR time series for Mount Marapi

Surface deformation (mm/ year)						
Years	2020	2021	2022	2023	2024	Average
Max Deformation	86.68	89.10	110.95	53.70	56.07	79.30
Min Deformation	-157.58	-67.19	-42.17	-61.93	-43.24	-74.42
Yearly Average	20.64	13.13	11.28	10.46	9.11	12.92
Stdv	-36.13	-5.59	14.48	-9.96	2.04	-7.03

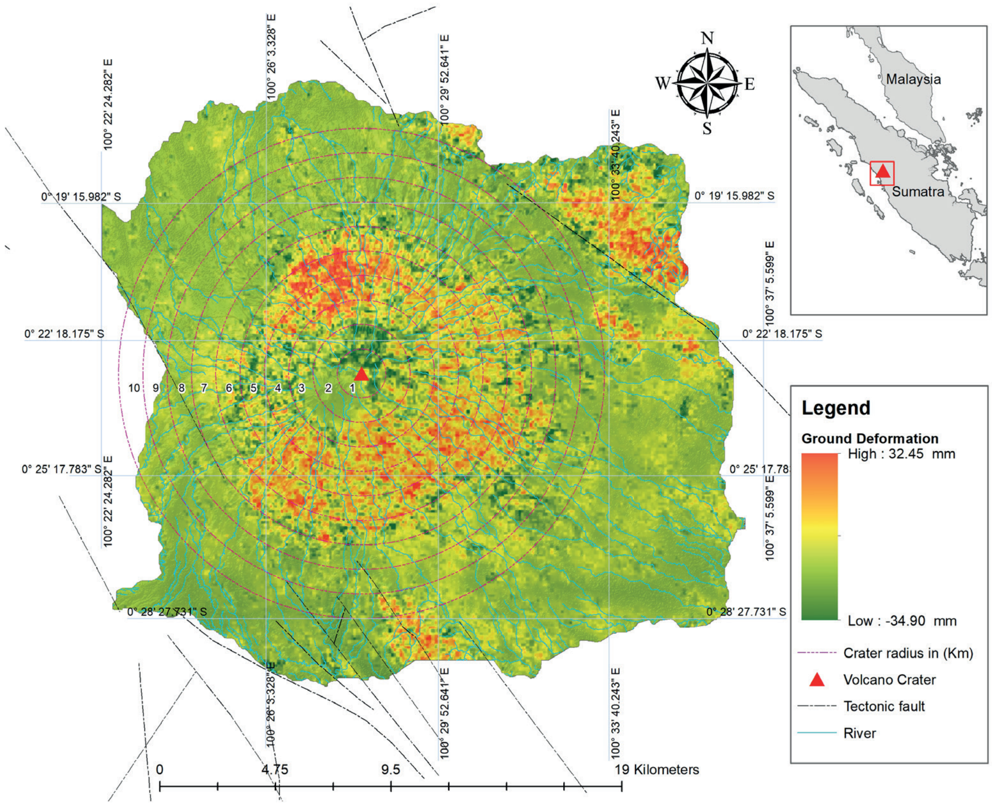

Fig. 2. Average annual deformation from 2020 to 2024

Table 3. Annual InSAR deformation at landslide occurrence points

Yearly deformation (mm/year)						
Years	2020	2021	2022	2023	2024	Average
Max Deformation	23.88	42.95	58.47	27.33	43.56	39.24
Min Deformation	-109.34	-36.72	24.51	34.27	-17.00	-44.37
Yearly Average	-51.19	0.71	12.16	-9.16	9.32	-7.63
Stdv	28.81	15.23	14.93	12.87	10.37	16.44

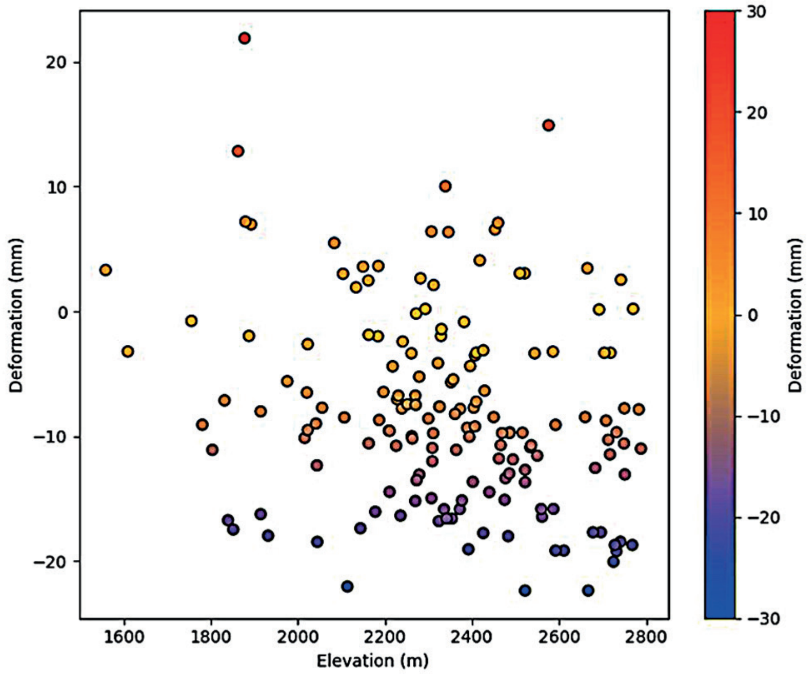

Fig. 3. Average annual deformation vs. elevation of each landslide point

Overall, all identified landslide sites exhibited indications of ground subsidence, with the average annual deformation showing negative displacement values (Fig. 3). Several locations recorded deformation values approaching −20 mm per year, which indicates persistent and periodic ground surface movement throughout the observation period of the deformation analysis. This sustained subsidence pattern suggests progressive slope instability that may culminate in peak subsidence events manifested as hillside collapse.

Landslide Susceptibility Zones

To predict landslides in the Mount Marapi area, a set of thematic spatial datasets is utilized, each representing factors that are interrelated and contribute to landslide susceptibility. The correlation matrix between these variables serves as a statistical tool to evaluate the strength and direction of their relationships, with correlation values ranging from −1 to 1. In this study, several key correlations were identified. Elevation exhibits a positive correlation with slope, river density, tectonic distance, and earthquake density, while showing a significant negative correlation with precipitation and land cover. Similarly, slope is positively correlated with elevation, river density, tectonic distance, and earthquake density, but negatively correlated with precipitation and land cover. River density also demonstrates positive correlations with elevation and slope but is negatively correlated with precipitation (See Fig 4). (Fig. 4).

From the correlation matrix, it can be inferred that precipitation exhibits a strong negative correlation with nearly all other variables, suggesting that higher rainfall is generally associated with lower values of factors such as elevation, slope, and related parameters. In contrast, both elevation and slope demonstrate significant positive correlations with variables including river density, tectonic distance, and earthquake density, indicating that topographic conditions may exert a considerable influence on these geomorphological and tectonic factors.

Each variable incorporated into the landslide prediction model contributes a distinct percentage to the overall predictive performance (see Fig. 5). Among these, three variables play a dominant role: slope, elevation, and precipitation. Slope emerges as the most influential factor, accounting for 37.99%, underscoring its critical role in determining landslide susceptibility in the Mount Marapi region. Elevation follows with a contribution of 17.28%, reflecting the influence of terrain height on soil stability and landslide potential, particularly within rugged and mountainous landscapes. Precipitation, contributing 13.76%, highlights the importance of rainfall as a triggering factor, where intense precipitation in upstream areas, especially on steep slopes, enhances soil saturation and markedly increases the likelihood of slope failure.

Other variables also demonstrate noteworthy contributions. Land cover accounts for 7.74%, indicating that vegetation type and land use significantly influence slope stability. Proximity to tectonic faults (5.92%) and

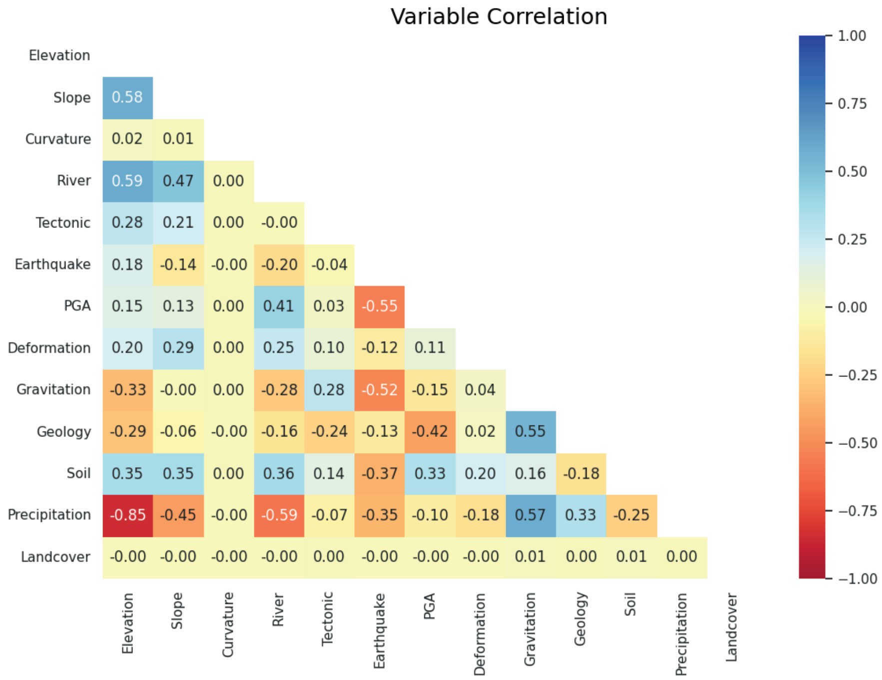

Fig. 4. The variable correlation for landslide prediction

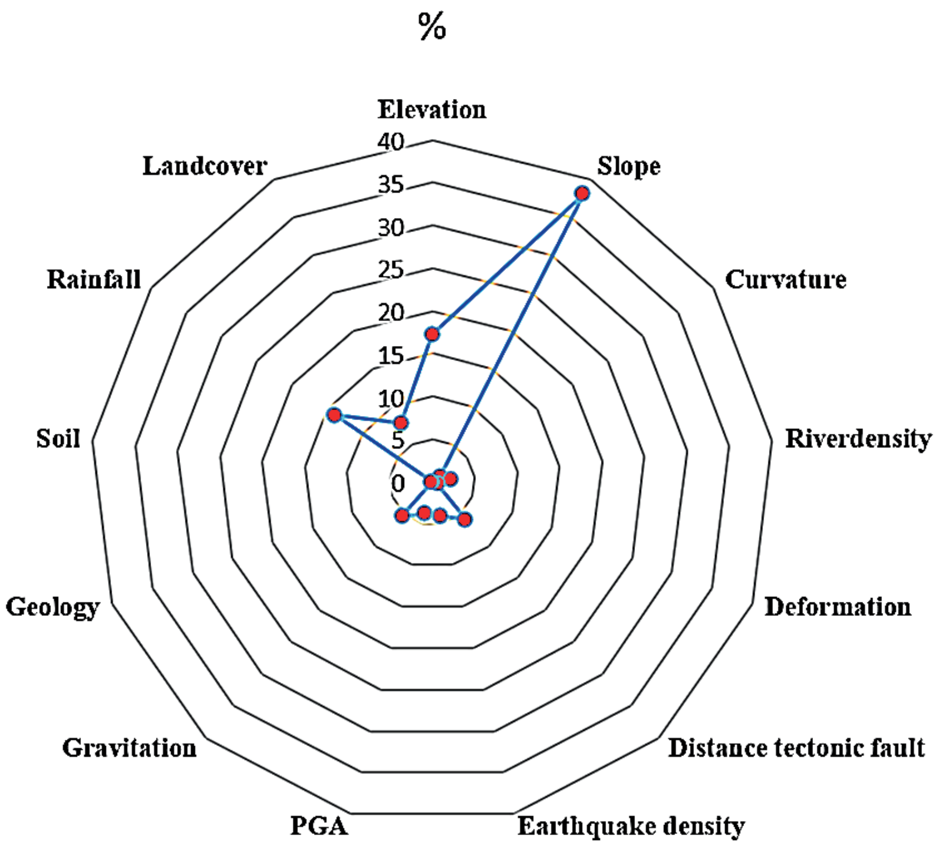

Fig. 5. The variable importance of landslide prediction in (%)

gravity-related factors (5.30%) further emphasize the role of geological activity and gravitational forces in slope instability. While variables such as land cover, tectonic fault distance, and gravity exert moderate effects, factors including curvature, ground deformation, and soil properties contribute minimally. Nevertheless, these parameters remain important in comprehensive landslide prediction analyses, as their combined interactions may still influence landslide occurrence.

After determining the optimal parameter configuration, the tuned parameters were subsequently applied to all machine learning algorithms implemented in this study, as summarized in the following Table 4. The predictive modeling results obtained through KNN, RF, SVM, and GB effectively delineated zones of potential landslide occurrence, assigning probability values ranging from 0 to 1. A value approaching 1 indicates a higher likelihood of landslide occurrence, thus providing a probabilistic basis for identifying and prioritizing high-risk areas (Fig. 6).

A performance comparison of multiple machine learning models based on various evaluation metrics is presented in Table 4. The models tested for landslide hazard prediction include KNN, RF, SVM, and GB. Among these, the Random Forest model demonstrated the highest overall performance, achieving an accuracy of 94.33% and a precision of 91.30%. These results indicate that Random Forest provides superior predictive accuracy with substantially lower error rates compared to the other models. KNN reached an accuracy of 92.45% and a precision of 84.61%. The Support Vector Classifier showed

Table 4. Model performance for landslide prediction

Multi-Machine Learning Approach	Accuracy %	Precision %	F1-Score	Recall	ROC
K-Nearest Neighbors	92.45	84.61	0.91	1	0.99
Random Forest	94.33	91.30	0.93	0.95	0.97
Support Vector Classifier	88.67	90.00	0.85	0.81	0.93
Gradient Boosting	90.56	90.47	0.88	0.86	0.97

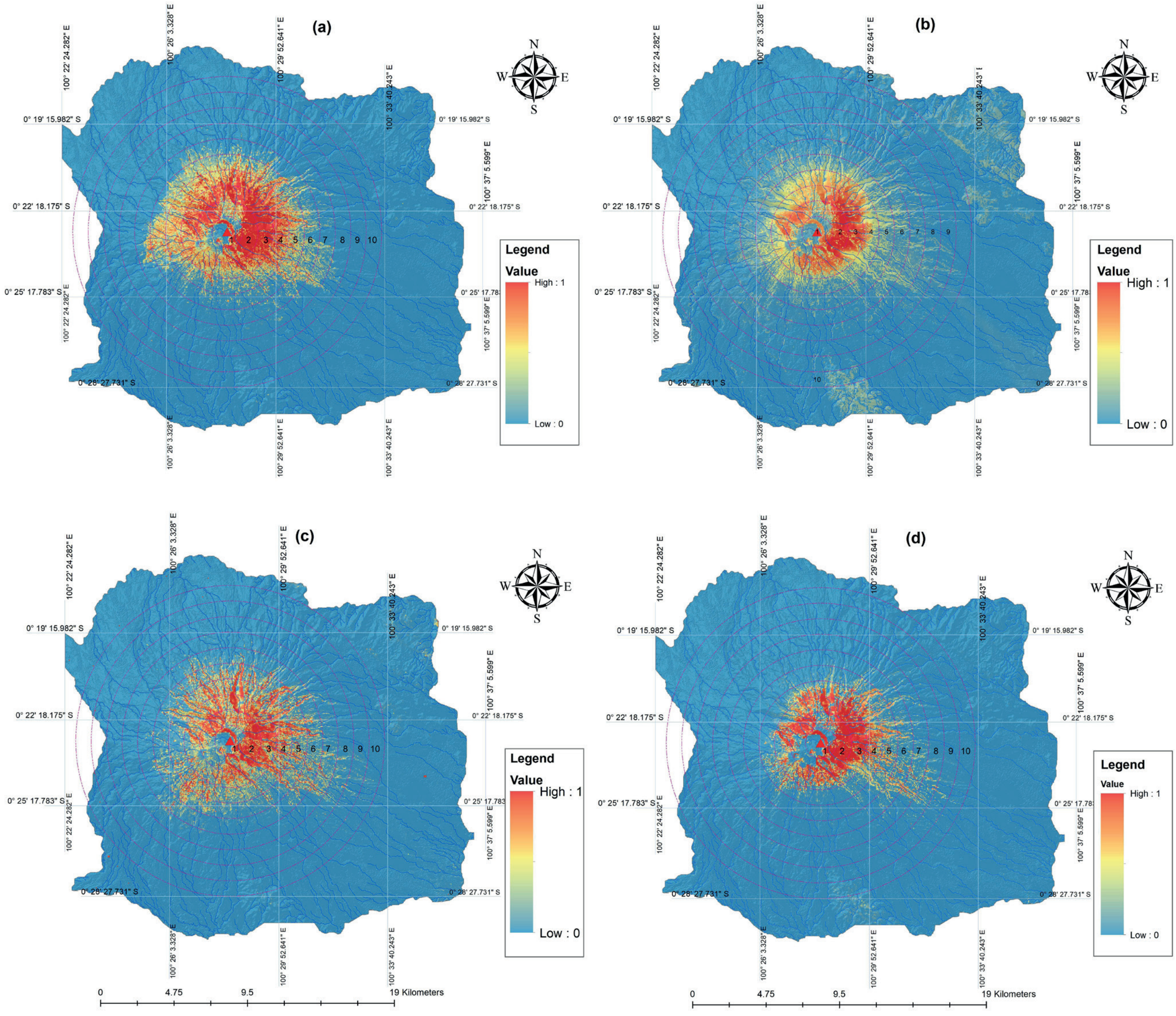

Fig. 6. The landslide probability around Mount Marapi: (a) K-Nearest Neighbors, (b) Random Forest, (c) Support Vector Classifier, and (d) Gradient Boosting.

the lowest predictive performance, further confirming its limited effectiveness for landslide prediction in the Mount Marapi region. The final landslide hazard prediction results generated by the Random Forest model are illustrated in Fig. 7.

Overall, the Mount Marapi region exhibits a higher probability of landslides in areas with steep slopes and elevated terrain, while relatively flat regions display a lower likelihood of occurrence. This spatial distribution delineates a hazard zone concentrated around the mountain's peak, suggesting that future landslides are likely to occur on multiple flanks of the volcano. The high-risk areas are characterized by elevated probability values compared to the lower foothill regions. This concentration of hazards near the peak is strongly correlated with the distribution of past landslide events, with 141 recorded incidents predominantly occurring on steep slopes and at higher elevations. In contrast, no landslide events have been documented in the foothill regions, where slopes are generally gentle or flat. The integration of landslide and non-landslide occurrence points plays a critical role in defining the spatial pattern of predicted susceptibility, as it directly influences the classification of spatial data across all contributing variables used in the analysis.

DISCUSSION

Deformation activity across the Mount Marapi region has generally exhibited a declining trend, although intermittent increases have been detected. On average, this indicates a net reduction in the mountain's surface elevation. This pattern is further corroborated by the progressive decrease in inflation rates, while deflation remains irregular and inconsistent, suggesting ongoing subsidence along the mountain slopes. The spatial distribution of deformation demonstrates a distinct configuration: moderate to steep slopes in proximity to the summit exhibit elevated inflation values, whereas the relatively flat foothill areas experience more stable and gradual deformation fluctuations. These displacement dynamics are strongly governed by geological processes, volcanic activity, and environmental factors (Seropian et al. 2021). Moreover, the accumulation of volcanic deposits and erosional processes further complicate the vertical deformation regime.

The high sensitivity and precision of Synthetic Aperture Radar (SAR) technology enable deformation mapping with millimeter-level accuracy (Hanif et al. 2024b). Time-series deformation analysis at landslide-prone zones of Mount Marapi reveals considerable fluctuations, with a gradual attenuation of peak inflation values. The persistence of surface deformation indicates the contribution of internal volcanic processes such as eruptive activity, hydrothermal pressure, and volcanic seismicity to ground instability (Hendra et al., 2019). Volcanic activity plays a critical role in destabilizing slopes, enhancing rock instability, particularly in steep terrains (Heap et al., 2021; Hanif et al. 2024c). The observed reduction in deformation values is also attributed to landslide events in several sectors, which substantially modify the mountain's surface morphology (Lu et al. 2024).

InSAR data derived from the LiCSAR platform provides a detailed characterization of these deformation patterns, offering valuable insights for landslide prediction. The integration of LiCSAR-derived deformation measurements with landslide susceptibility models substantially strengthens hazard assessments. Deformation data functions as a crucial explanatory variable, capturing land surface dynamics that directly influence slope stability. Areas experiencing pronounced subsidence or localized uplift can thus be delineated as potential high-risk zones.

By integrating ground deformation data with additional predictive variables, a highly accurate landslide hazard map of Mount Marapi has been developed. Among the factors influencing landslide susceptibility within the Random Forest framework, slope, elevation, and precipitation exert the strongest control. Steep topography in high-elevation zones near the crater, combined with episodes of intense rainfall, markedly elevates the probability of slope failure. Mount Marapi's tropical montane environment further amplifies this susceptibility: dense forest cover,

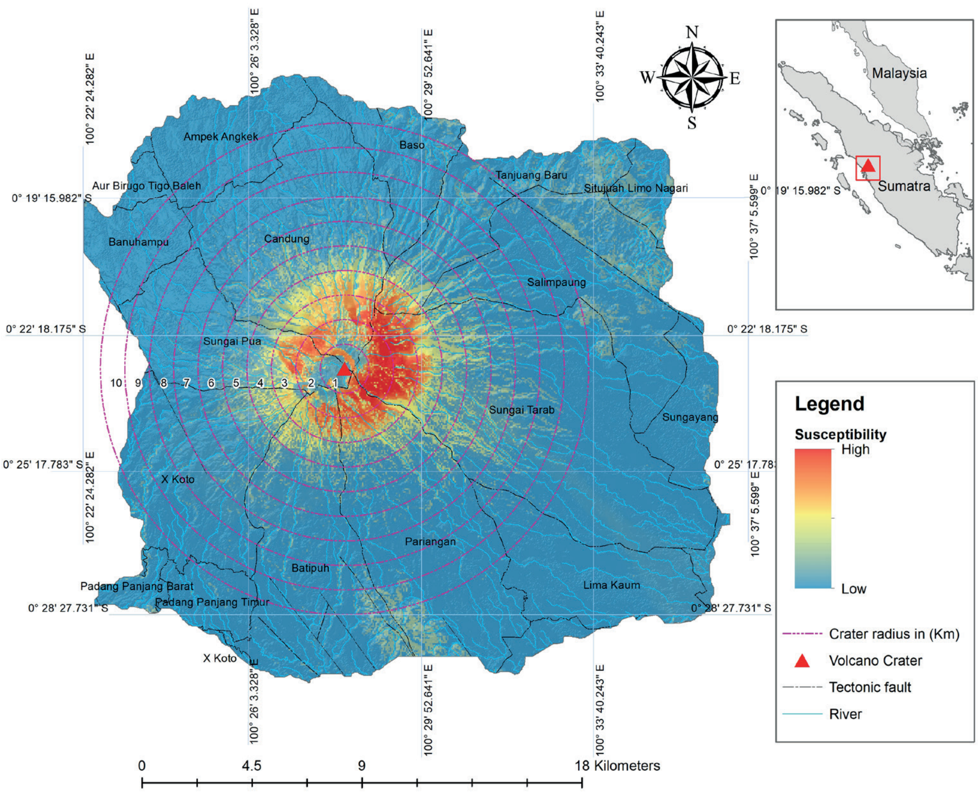

Fig. 7. The landslide susceptibility on Mount Marapi

high atmospheric humidity, and abundant precipitation, coupled with a relatively thin soil mantle overlying young volcanic deposits, predispose slopes to rapid soil saturation. The interplay of geomorphological, geological, and environmental variables collectively exacerbates slope instability, as these processes are interdependent and act synergistically to intensify landslide risk (Hussain et al. 2022; Tehraini et al. 2022).

Ground deformation signals derived from InSAR are strongly correlated with historical landslide events on Mount Marapi. Localized deformation on steep slopes, particularly in sectors experiencing uplift (inflation) or subsidence, contributes directly to terrain instability. Inflation increases internal stress within the volcanic edifice, whereas subsidence produces compaction and surface fissures, both of which promote mass movement. These processes are further accelerated under conditions of heavy rainfall, as infiltration elevates pore-water pressure and reduces soil shear strength, thereby heightening the likelihood of slope failure.

Among the various machine learning models applied for landslide prediction in the Mount Marapi region, the RF algorithm demonstrated superior performance in handling the spatial datasets, which consist of complex combinations of categorical and continuous variables. The RF model effectively delineated landslide susceptibility zones in an area with a documented history of landslide occurrences. This superiority arises from the ability of decision trees to capture non-linear relationships and construct intricate decision boundaries without relying on distributional assumptions (Li et al. 2024). Furthermore, the ensemble nature of RF, which aggregates multiple trees trained on randomized subsets of data and predictor variables, enhances model stability, reduces sensitivity to noise, and captures variable interactions more effectively, thereby producing results that are both accurate and generalizable (Lu et al. 2024).

The spatial distribution demonstrates a strong correlation between historical landslide occurrences and predictions generated by the RF model. Landslide events in mountainous terrain are not randomly distributed but instead form clusters on steep slopes at higher elevations, particularly around mountain peaks. This distribution corresponds with key controlling factors such as slope gradient, intense rainfall, and young volcanic lithology, which are spatially concentrated in specific zones. The Random Forest predictions reproduce this clustering tendency, with high probabilities of landslides along the mountain flanks and low probabilities across flatter terrain at the mountain base, thereby reinforcing confidence in the model's reliability. Nevertheless, as noted by Wen et al. (2024), strong spatial autocorrelation introduces the risk of spatial bias, whereby the model may become overly dependent on clustered historical patterns. As a result, although effective in identifying susceptibility zones resembling past events, the model may exhibit reduced sensitivity in detecting newly emerging vulnerable areas outside established clusters (Kanwar et al. 2025).

Furthermore, the results of landslide susceptibility zoning are strongly influenced by multiple factors, particularly Mount Marapi's tectonic position within an active subduction zone, which makes the region highly prone to seismic disturbances and highlights the importance of geologically based hazard assessment. Several geophysical conditions shape landslide dynamics in the area, including active tectonic plate convergence, seismic history, peak ground acceleration, and geomagnetic variations. The Semangko Fault, an active strike-slip system undergoing continuous displacement, generates vertical and lateral deformation that frequently triggers tectonic earthquakes. Seismic shaking from these events reduces slope shear resistance and destabilizes surface materials, with its magnitude commonly quantified by peak ground acceleration (Chen et al. 2011; Xie et al. 2022). When coupled with intense rainfall, this seismic instability creates critical conditions for slope failure, particularly in steep and saturated terrains (Huang et al. 2024). Although secondary factors such as curvature, deformation rate, and soil composition exert relatively weaker direct effects, their cumulative and interdependent roles underscore the necessity of integrating them into a comprehensive, multi-variable landslide prediction framework.

The RF model produced accurate predictions; several biases may limit its reliability. Spatial bias arises because landslide events around Mount Marapi are concentrated in specific areas, while temporal bias reflects the restricted availability of historical inventories and the omission of extreme rainfall periods. Such limitations highlight that general computational principles cannot be directly applied to spatial data that represent real-world Earth surface processes, where geomorphic processes are highly context-dependent. For instance, steep slopes under intense rainfall are highly susceptible to landslides, whereas flat terrain may instead experience flooding due to poor drainage. This suggests that landslide prediction cannot rely solely on numerical or machine learning approaches (Chandran and Roy, 2024). A more robust framework requires integrating computational methods with fundamental geographical perspectives, drawing on spatial, environmental, and regional approaches for spatial data interpretation.

CONCLUSIONS

Deformation analysis using InSAR and landslide prediction through machine learning models provide complementary perspectives on land surface dynamics at Mount Marapi. The deformation trend indicates a gradual decline accompanied by periodic fluctuations, characterized by decreasing inflation rates and irregular deflation patterns, particularly along steep slopes near the summit. These deformation signals are primarily governed by volcanic activity, hydrothermal pressurization, and seismic events, all of which progressively destabilize rock masses and increase the likelihood of slope failure.

The application of ensemble and multi-machine learning approaches has further enhanced the predictive performance of landslide susceptibility models, with the Random Forest algorithm demonstrating the highest classification accuracy. Landslide occurrence is predominantly controlled by three critical factors: slope gradient, elevation, and rainfall intensity. High-altitude, steeply inclined slopes are especially vulnerable, particularly during periods of heavy precipitation. Tropical humidity, combined with thin soil layers, further amplifies instability, while seismic activity acts as a compounding trigger for slope failure. Additional conditioning factors including slope curvature, surface deformation anomalies, and soil type exert more subtle, yet still significant, influences on model predictions.

Deformation measurements capture surface displacements that directly affect slope stability, whereas machine learning-based prediction models delineate areas with heightened susceptibility to landslides. A holistic, integrated framework that synthesizes deformation monitoring with machine learning-based susceptibility mapping is therefore essential for comprehensive hazard assessment and effective risk mitigation. Importantly, the landslide prediction framework employed in this study is based on open-source datasets, underscoring its scalability and potential applicability to diverse geomorphic and volcanic environments worldwide.

Despite its contributions, this research is subject to several limitations. The spatial resolution of DInSAR data from LiCSAR constrains its ability to detect small-scale deformation, particularly in areas with dense vegetation cover and steep topography. Second, the modeling framework incorporates a limited set of variables, which may not fully capture the complexity of field conditions. In addition, the predictive models rely heavily on historical landslide inventories and satellite-derived rainfall data, the latter of which may underrepresent localized extreme precipitation events that act as critical landslide triggers. Another important limitation lies in the potential overestimation of model accuracy due to the absence of extensive field validation, combined with the inherently static nature of the model, which does not account for rapidly evolving geomorphological processes. For these reasons, the prediction outputs should be interpreted with caution, particularly in highly dynamic volcanic and mountainous environments where slope stability conditions can change abruptly.

REFERENCES

Abu El-Magd S.A., Ali S.A. and Pham Q.B. (2021). Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. Earth Science Informatics, 14(3), 1227–1243, DOI: 10.1007/s12145-021-00653-y.

Adnan M.S.G., Rahman M.S., Ahmed N., Ahmed B., Rabbi M.F. and Rahman R.M. (2020). Improving spatial agreement in machine learning-based landslide susceptibility mapping. Remote Sensing, 12(20), 3347, DOI: 10.3390/rs12203347.

Chandran A. and Roy P. (2024). Applications of geographical information system and spatial analysis in Indian health research: a systematic review. BMC Health Services Research, 24(1), 1448, DOI: 10.1186/s12913-024-11837-9.

Chowdhury M.S., Rahman M.N., Sheikh M.S., Sayeid M.A., Mahmud K.H. and Hafsa B. (2024). GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. Heliyon, 10(1), e23424, DOI: 10.1016/j.heliyon.2023.e23424.

Chen H., Lin G.W., Lu M.H., Shih T.Y., Horng M.J., Wu S.J. and Chuang B. (2011). Effects of topography, lithology, rainfall, and earthquake on landslide and sediment discharge in mountain catchments of southeastern Taiwan. Geomorphology, 133(3–4), 132–142, DOI: 10.1016/j.geomorph.2010.12.031.

Demertzis K., Iliadis L. and Pimenidis E. (2021). Geo-AI to aid disaster response by memory-augmented deep reservoir computing. Integrated Computer-Aided Engineering, 28(4), 383–398, DOI: 10.3233/ICA-210816.

Hanif M., Apichontrakul S. and Mokhtar E.S. (2024a). Clustering and forecasting vertical ground deformation of Sinabung Volcano based on deformation dataset and Holt-Winters exponential smoothing method. In: International Conference of Geoscience and Remote Sensing Technology. Singapore: Springer Nature Singapore, 50–63, DOI: 10.1007/978-981-95-3075-5_5.

Hanif M., Apichontrakul S. and Razi P. (2024b). Surface deformation monitoring and forecasting of Sinabung volcano using interferometry synthetic aperture radar and forest-based algorithm. Remote Sensing Applications: Society and Environment, 36, 101288, DOI: 10.1016/j.rsase.2024.101288.

Hanif M., Apichontrakul S. and Razi P. (2024c). Multi-temporal InSAR analysis for monitoring the ground deformation of Mount Sinabung. Asia-Pacific Journal of Science and Technology, 29(5), 1–9, DOI: 10.14456/apst.2024.83.

Heap M., Tobias B., Albert G.H., Stephan K. and Amy R. (2021). Hydrothermal alteration can result in pore pressurization and volcano instability. Geology, 49(11), 1348–1352, DOI: 10.1130/G49063.1.

Hendra G., Surono A.B., Kristianto O.P., Wendy M.C., John P. and Masato I. (2019). Overview of the eruptions of Sinabung Volcano, 2010 and 2013–present and details of the 2013 phreatomagmatic phase. Journal of Volcanology and Geothermal Research, DOI: 10.1016/j.jvolgeores.2017.08.005.

Huang Y., Xu C., He X., Cheng J., Huang Y., Wu L. and Xu X. (2024). Distribution characteristics and cumulative effects of landslides triggered by multiple moderate-magnitude earthquakes: a case study of the comprehensive seismic impact area in Yibin, Sichuan, China. Landslides, 21(12), 2927–2943, DOI: 10.1007/s10346-024-01626-7.

Huang W., Ding M., Li Z., Zhuang J., Yang J., Li X., Meng L.E., Zhang H. and Dong Y. (2022). An efficient user-friendly integration tool for landslide susceptibility mapping based on support vector machines: SVM-LSM toolbox. Remote Sensing, 14(14), 3408, DOI: 10.3390/rs14143408.

Hussain M.A., Chen Z., Zheng Y., Shoaib M., Shah S.U., Ali N. and Afzal Z. (2022). Landslide susceptibility mapping using machine learning algorithm validated by persistent scatterer InSAR technique. Sensors, 22(9), 3119, DOI: 10.3390/s22093119.

Kanwar M., Pokharel B. and Lim S. (2025). A new random forest method for landslide susceptibility mapping using hyperparameter optimization and grid search techniques. International Journal of Environmental Science and Technology, 1–16, DOI: 10.1007/s13762-024-06310-3.

Kavzoglu T. and Teke A. (2022). Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). Arabian Journal for Science and Engineering, 47(6), 7367–7385, DOI: 10.1007/s13369-022-06560-8.

Lazecký M., Spaans K., González P.J., Maghsoudi Y., Morishita Y., Albino F., Elliott J., Greenall N., Hatton E.L., Hooper A., Juncu D., McDougall A., Walters R.J., Watson C., Weiss J.R. and Wright T. (2020). LiCSAR: An automatic InSAR tool for measuring and monitoring tectonic and volcanic activity. Remote Sensing, 12, 424, DOI: 10.3390/rs12030424.

Le Breton M., Bontemps N., Guillemot A., Baillet L. and Larose É. (2021). Landslide monitoring using seismic ambient noise correlation: challenges and applications. Earth-Science Reviews, 216, 103518, DOI: 10.1016/j.earscirev.2021.103518.

Li M., Wang H., Chen J. and Zheng K. (2024). Assessing landslide susceptibility based on the random forest model and multi-source heterogeneous data. Ecological Indicators, 158, 111600, DOI: 10.1016/j.ecolind.2024.111600.

Lu M., Tay L.T. and Mohamad-Saleh J. (2024). Landslide susceptibility analysis using random forest model with SMOTE-ENN resampling algorithm. Geomatics, Natural Hazards and Risk, 15(1), 2314565, DOI: 10.1080/19475705.2024.2314565.

Lu Q.O. and Lee C.C. (2025). Innovative Geo-AI model: an enhancement of outdoor PM estimations based on land use and outdoor environmental factors in a highly polluted area. Chemosphere, 373, 144178, DOI: 10.1016/j.chemosphere.2023.144178.

Mao Y., Li Y., Teng F., Sabonchi A.K., Azarafza M. and Zhang M. (2024). Utilizing hybrid machine learning and soft computing techniques for landslide susceptibility mapping in a drainage basin. Water, 16(3), 380, DOI: 10.3390/w16030380.

Mondini A.C., Santangelo M., Rocchetti M., Rossetto E., Manconi A. and Monserrat O. (2019). Sentinel-1 SAR amplitude imagery for rapid landslide detection. Remote Sensing, 11, 760, DOI: 10.3390/rs11060760.

Morishita Y., Lazecky M., Wright T.J., Weiss J.R., Elliott J.R. and Hooper A. (2020). LiCSBAS: An open-source InSAR time series analysis package integrated with the LiCSAR automated Sentinel-1 InSAR processor. Remote Sensing, 12, 424, DOI: 10.3390/rs12030424.

Narny Y., Putri F.R., Cribb R., Denny D., Sari D.P., Sari E.S., Ismail F.A., Anwar K., Nasta A.A. and Yuliardi Y. (2025). Lahar floods in West Sumatra during the Dutch colonial era: historical records and societal impact. In: E3S Web of Conferences, 604, 11002, DOI: 10.1051/e3sconf/202560411002.

Pei T. and Qiu T. (2024). Landslide susceptibility mapping using physics-guided machine learning: a case study of a debris flow event in the Colorado Front Range. Acta Geotechnica, 19(10), 6617–6641, DOI: 10.1007/s11440-024-01440-4.

Sahin E.K. (2022). Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. Geocarto International, 37(9), 2441–2465, DOI: 10.1080/10106049.2020.1831623.

Seropian G., Kennedy B.M. and Walter T.R. (2021). A review framework of how earthquakes trigger volcanic eruptions. Nature Communications, 12, 1004, DOI: 10.1038/s41467-021-21166-8.

Su Y., Chen Y., Lai X., Huang S., Lin C. and Xie X. (2024). Feature adaptation for landslide susceptibility assessment in "no sample" areas. Gondwana Research, 131, 1–17, DOI: 10.1016/j.gr.2024.03.002.

Tehrani F.S., Calvello M., Liu Z., Zhang L. and Lacasse S. (2022). Machine learning and landslide studies: recent advances and applications. Natural Hazards, 114(2), 1197–1245, DOI: 10.1007/s11069-022-05423-7.

Triyatno, Anwar S. and Fabriandi R.L. (2024). GIS based on analysis of earthquake hazard level using the PGA in Siberut–Mentawai Islands Regency for sustainability in disaster mitigation. Journal of Sustainability Science and Management, 19(12), 107–119, DOI: 10.46754/jssm.2024.12.007.

Wen H., Li J., Liao M., Di M., Hu J. and Liu B. (2025). A hybrid-optimized random forest interpretable model for debris flow susceptibility by prior model-based negative sampling. Advances in Space Research, 76(1), 202–220, DOI: 10.1016/j.asr.2025.04.055.

Wu B., Shi Z., Zheng H., Peng M. and Meng S. (2024). Impact of sampling for landslide susceptibility assessment using interpretable machine learning models. Bulletin of Engineering Geology and the Environment, 83(11), 1–19, DOI: 10.1007/s10064-024-02008-w.

Wulansari T., Wulandini D.R. and Harahap E.F. (2024). Analysis of the impact of flash floods and cold lava floods on Mount Marapi, West Sumatra 2024. In: Proceeding of International Conference on Digital, Social, and Science, 1(1), 224–230, DOI: 10.62201/icodss.v1i01.166.

Xia D., Tang H., Sun S., Tang C. and Zhang B. (2022). Landslide susceptibility mapping based on the germinal center optimization algorithm and support vector classification. Remote Sensing, 14(11), 2707, DOI: 10.3390/rs14112707.

Xie J., Coulthard T.J. and McLelland S.J. (2022). Modelling the impact of seismic triggered landslide location on basin sediment yield, dynamics and connectivity. Geomorphology, 398, 108029, DOI: 10.1016/j.geomorph.2021.108029.

Xu Z., Che A. and Zhou H. (2024). Seismic landslide susceptibility assessment using principal component analysis and support vector machine. Scientific Reports, 14(1), 3734, DOI: 10.21203/rs.3.rs-761260/v1.

Zhang S., Wang Y. and Wu G. (2022). Earthquake-induced landslide susceptibility assessment using a novel model based on gradient boosting machine learning and class balancing methods. Remote Sensing, 14(23), 5945, DOI: 10.3390/rs14235945.