



Article

InSAR Integrated Machine Learning Approach for Landslide Susceptibility Mapping in California

Divya Sekhar Vaka , Vishnuvardhan Reddy Yaragunda, Skevi Perdikou * and Alexandra Papanicolaou

GEOFEM, 1080 Nicosia, Cyprus; sekhar.vaka@geofem.com (D.S.V.); vishnu.yaragunda@geofem.com (V.R.Y.); alexandra.papanicolaou@geofem.com (A.P.)

* Correspondence: skevi.perdikou@geofem.com

Abstract: Landslides pose significant threats to life and property, particularly in mountainous regions. To address this, this study develops a landslide susceptibility model integrating Earth Observation (EO) data, historical data, and Multi-Temporal Interferometric Synthetic Aperture Radar (MT-InSAR) ground movement results. The model categorizes areas into four susceptibility classes (from Class 1 to Class 4) using a multi-class classification approach. Results indicate that the Xtreme Gradient Boosting (XGB) model effectively predicts landslide susceptibility with area under the curve (AUC) values ranging from 0.93 to 0.97, with high accuracy of 0.89 and a balanced performance across different susceptibility classes. The integration of MT-InSAR data enhances the model's ability to capture dynamic ground movement and improves landslide mapping. The landslide susceptibility map generated by the XGB model indicates high susceptibility along the Pacific coast. The optimal model was validated against 272 historical landslide occurrences, with predictions distributed as follows: 68 occurrences (25%) in Class 1, 142 occurrences (52%) in Class 2, 58 occurrences (21.5%) in Class 3, and 4 occurrences (1.5%) in Class 4. This study highlights the importance of considering temporal changes in environmental conditions such as precipitation, distance to streams, and changes in vegetation for accurate landslide susceptibility assessment.

Keywords: classification; XGBoost; random forest; PSInSAR; SBAS; MT-InSAR; slope stability



Citation: Vaka, D.S.; Yaragunda, V.R.; Perdikou, S.; Papanicolaou, A. InSAR Integrated Machine Learning Approach for Landslide Susceptibility Mapping in California. *Remote Sens.* **2024**, *16*, 3574. <https://doi.org/10.3390/rs16193574>

Academic Editor: Matteo Albano

Received: 27 July 2024

Revised: 20 September 2024

Accepted: 23 September 2024

Published: 25 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslides are major natural disasters globally, primarily driven by gravitational forces along with factors such as earthquakes, precipitation, and human activities [1]. Landslides are caused by external factors that increase shear stress (e.g., unloading the slope toe, loading the slope crest, shocks and vibrations, changes in the water regime, and geometrical changes) and internal factors that decrease shearing resistance (e.g., weathering, progressive failure, and seepage erosion) [2]. Each year, landslides, whether triggered by human activities or natural events, cause significant economic damage and loss of life [3]. In particular, the frequent occurrence of landslides along roads and cut slopes in mountainous regions poses a substantial threat to residents in these areas [4]. Between 1998 and 2017, an estimated 4.8 million people were affected, and over 18,000 deaths occurred due to landslides, with low-income populations being the most affected [5]. It is believed that recognizing and addressing the issue before landslide events occur could prevent at least 90% of the losses, highlighting the importance of preventive measures [6].

In this scenario, continuous monitoring of landslide-prone areas is crucial. Monitoring includes tracking both short-term rapid deformations and long-term slow-moving deformations on the slopes. Traditional landslide susceptibility mapping is a time-consuming and complex process that involves extensive field data collection and analysis. Regularly conducting field-based activities to monitor changes and signs of recent landslides, especially those triggered by events like rainfall, is impractical, particularly when mapping large areas [5].

Over the past two decades, advanced techniques have emerged for detecting deformation, such as leveling, global positioning systems (GPS), and geotechnical methods. However, these approaches often have significant drawbacks, including high costs, time-consuming procedures, and limited coverage, offering continuous monitoring for only a small portion of landslide-prone areas [7].

In this context, landslide susceptibility mapping (LSM) using Earth Observation (EO) data is gaining popularity. These methods offer several advantages over conventional techniques, including low cost, repeatable and efficient procedures, data-driven scalability, and wider coverage with a good revisit cycle from remote sensing satellites. The direct use of satellite remote sensing imagery and its derivatives is widely employed in creating landslide susceptibility maps (LSM) over extensive areas.

For example, factors such as slope, aspect, curvature, roughness, and stream network, derived from digital elevation models (DEM), are widely used in landslide susceptibility mapping. In general, slope-derived factors are crucial indicators of landslide occurrence. According to Popescu [2], the diverse range of slope movements indicates various conditions leading to slope instability and the processes driving such movements. Kirschbaum et al. [8] combined remote sensing derivatives with in situ data and used six parameters—slope, soil type, soil texture, elevation, land cover, and drainage density—to determine the contribution of each variable class to landslide susceptibility estimation. A few studies [9–12] have explored the relationship between slope angle and aspect in landslide susceptibility, concluding that a slope angle between 20° and 40° implies the highest susceptibility to landslides. Combining the aforementioned factors with other remote sensing-derived layers such as land cover, precipitation, road networks, and vegetation indices can enhance the accuracy of landslide susceptibility maps. Evans et al. [13] identified key preliminary landslide pre-conditioning factors, including geology, topography, vegetation cover, tectonic activity, and quaternary history. They also recognized rainfall, loading/unloading, water level changes, and earthquakes as significant causal factors contributing to landslide occurrences.

In recent years, Interferometric Synthetic Aperture Radar (InSAR) has become widely utilized for mapping geohazards on large scales, such as landslides [14,15], sinkholes [16], earthquake deformations [17,18], etc. The increased availability of Synthetic Aperture Radar (SAR) imagery and advancements in various techniques have significantly enhanced the popularity of InSAR data in landslide mapping. Multi-Temporal InSAR (MT-InSAR) methods, such as Persistent Scatter Interferometry (PSI) [19] and Small Baseline Subset (SBAS) [20,21], are commonly employed for landslide mapping and monitoring [22–28]. These technological advancements using InSAR techniques aid in developing landslide inventory data, characterizing landslides, and quantifying their impacts [22].

The current assessment of landslide susceptibility typically relies on models built using static factors [29–31], neglecting dynamic features such as ground deformation [27]. This omission can lead to inaccuracies in identifying certain landslide-prone areas. Integrating MT-InSAR results into machine learning (ML)-based landslide susceptibility models holds great promise. By incorporating ground movement data derived from MT-InSAR methods, these models can account for the dynamic behavior of landslides [25–28]. ML models are known for their excellent performance and efficient modeling process in capturing the relationship between various factors and landslides, making them widely used in landslide susceptibility mapping. Recently, ML models such as logistic regression [28,32], support vector machines [27,32], random forests [25,27,28,32], naive Bayes [25], and Xtreme gradient boosting [25,28,32] have been widely applied in LSM. Comparative analysis of ML-based LSM and MT-InSAR deformation results shows that areas with higher deformation points correspond to higher susceptibility levels, while areas with lower susceptibility points are more stable [33].

Handwerker et al. [34] use a novel InSAR detection technique that applies both local and regional filters to reduce background noise and highlight specific deformations, like landslides. The results from this study have identified hundreds of slow-moving landslides

in California, with particularly high concentrations along the Big Sur coast, the central San Andreas Fault, and the Eel River area—all known for frequent landslides. By comparing the displacement time series of numerous landslides with local precipitation data, the authors observed a direct correlation between cumulative rainfall and landslide movement. Kang et al. [35] analyzed Sentinel-1 ascending and descending InSAR datasets from 2014 to 2016 and identified many landslides along the Highway 50 corridor in California moving at rates of less than 10 cm/year. Notably, peak landslide deformation often occurs in the dry season (May to October) due to a delay in precipitation infiltration. Cohen-Waeber et al. observed that active slope deformation across the Lawrence Berkeley National Laboratory (LBNL) site and the San Francisco East Bay hills (Berkeley Hills) results from various static and dynamic conditions [36,37]. A review [36] of three separate InSAR time series analyses of the Berkeley Hills—using ERS-1/2 data from 1992 to 2001 [38], data from 2001 to 2006 [39], and data from 2009 to 2011 [40]—shows remarkable consistency. These independent studies found that precipitation-related displacement did not occur immediately, but with lag times of up to 1 to 3 months. In another study, Cohen-Waeber et al. [41] utilized TerraSAR-X satellite images from 2009 to 2014 and a proprietary MT-InSAR algorithm to generate a highly detailed time series of ground deformation for the San Francisco East Bay Hills. The independent and principal component analyses of this time series uncover four distinct spatial and temporal deformation patterns in the area around the Blakemont landslide in California. Two of these components identify continuous landslide movement and deformation driven by precipitation-related pore pressure changes, influenced by annual seasonal cycles and multiyear drought conditions. The remaining two components, representing more widespread seasonal deformation, distinguish between precipitation-induced soil swelling and annual variations potentially linked to groundwater level fluctuations and thermal expansion of structures.

The findings highlight the effectiveness of MT-InSAR in detecting slow landslide movements in challenging terrains. While MT-InSAR technology has been used in various studies to identify landslides, its application in large-scale LSM across extensive areas in California has been relatively underexplored. Additionally, existing ML-based LSM often incorporates MT-InSAR velocity estimates as input weights but has not extensively investigated the direct integration of ground movement results with other landslide causative factors in the ML model. Our current research aims to address these gaps by focusing on (1) large-scale landslide susceptibility mapping using MT-InSAR and (2) the direct integration of MT-InSAR data alongside other landslide-causative parameters in ML models.

Therefore, the present study aims to develop and evaluate a landslide susceptibility model that integrates conventional EO data derivatives (e.g., slope, aspect, and NDVI), historical data (e.g., rainfall and geology), and MT-InSAR ground movement results. The objective is to create a multi-class classification model utilizing these diverse datasets to categorize geographical areas into four distinct susceptibility levels: no susceptibility at this scale and with the available information, low susceptibility, medium susceptibility, and high susceptibility.

2. Study Area

The study area presented in this paper is located in the state of California, as depicted in Figure 1. Situated on the west coast of the United States, California features a diverse geography that includes rich valleys, coastal areas, deserts, and mountain ranges. With its large population and significant economic impact, California is one of the most prominent states in the country. The study area is characterized by complex topography, seismic activities, and seasonal weather patterns, making it prone to landslides. The combination of steep slopes, heavy rainfall, and seismic events increases the risk of slope instability and potential landslides [24].

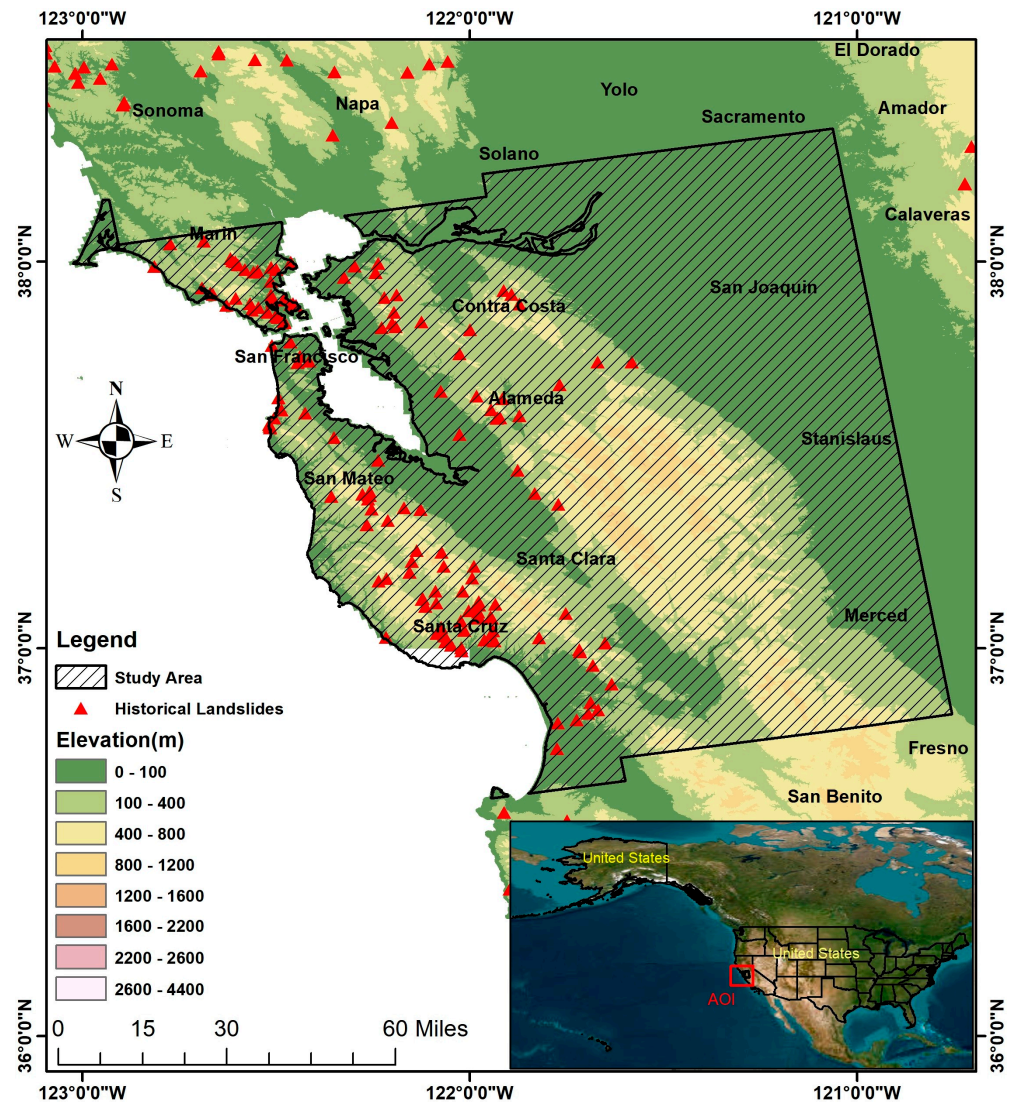


Figure 1. Study area map of the San Francisco region, featuring a United States Geological Survey (USGS) 3D Elevation Program (3DEP) Digital Elevation Model as the background. Historical landslide locations are denoted with red triangles (with 272 landslides falling within the study area AOI), and the extent of the Sentinel—1 Synthetic Aperture Radar (SAR) image is outlined with a black polygon.

According to the landslide inventory catalog [25], the San Francisco area experiences frequent landslides. For these reasons, San Francisco has been selected as the study area for the landslide susceptibility analysis.

The geology of the study area (Figure 2) includes sedimentary rocks such as sandstones, limestone, and shales, found in the coastal ranges and central valleys. Volcanic rocks like andesite and basalt are present in the northern part of the state. The San Andreas Fault, a major geological feature, consists of fractured and deformed rocks, contributing to frequent seismic activity in the region [42].

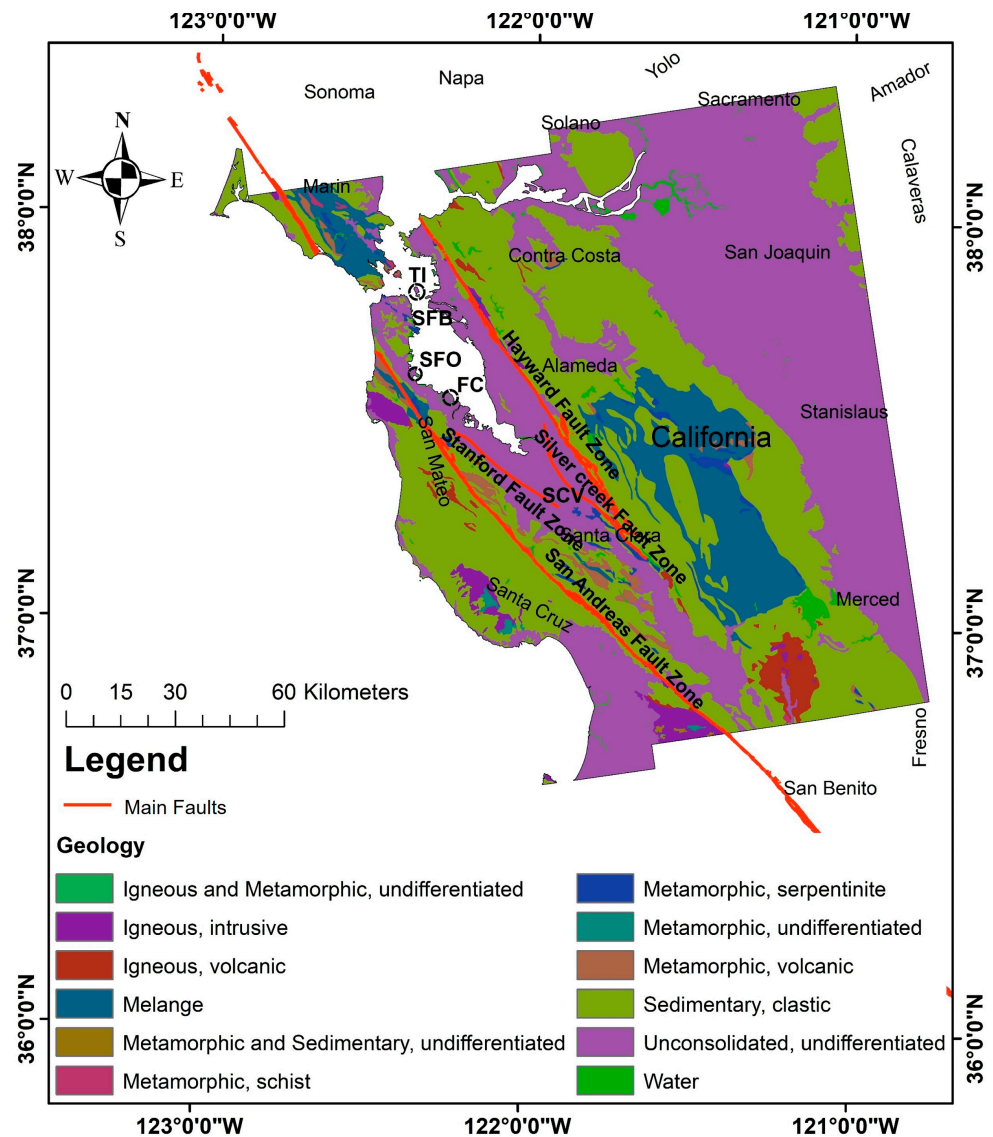


Figure 2. The geological profile of the study area was obtained from the USGS website [43], based on a 1:250,000 scale digitized map. Major faults overlaid on the map are retrieved from the California State Geoportal (<https://gis.data.ca.gov/>, accessed on 11 August 2023).

3. Data Collection and Preparation

The dataset comprised nine Geographic Information System (GIS) layers derived from DEM and historical geospatial data as predictor variables. The MT-InSAR results along with slope values are used to classify the target variable into four different classes as described in Section 4. The selected features for our models included both numeric and categorical variables such as slope, aspect, distance from a stream, annual precipitation, vegetation, curvature, flow direction, soil type, and geology. Each feature type required specific handling to ensure accurate analysis and model integration. These features underwent data processing, data cleaning to remove records with missing values, and significant data transformations, such as categorization and dummy variable encoding, to facilitate their effective utilization by the machine learning models. Each factor is described in the following sections:

3.1. GIS Layers

The GIS layers used as an input to the model include slope, aspect, curvature, flow direction, distance to a stream, rainfall, vegetation index, soil type, and geology (Figure 3).

The first five GIS layers are developed from the USGS 3DEP 10 m spatial resolution DEM using ArcGIS software v10.5. Rainfall, soil type, and geology layers are developed from historical geospatial data, whereas the vegetation index (NDVI) is created by processing Sentinel-2 images in Google Earth Engine. Incorporating historical data on rainfall, soil, and geological conditions is crucial for precise assessments, as it provides valuable insights into the factors contributing to slope failures. Understanding the variability of these factors and their influence on landslide occurrences enhances the ability to predict susceptibility more effectively. Out of these historical parameters, rainfall is a primary trigger: intense or prolonged precipitation increases soil moisture and ultimately decreases soil shear strength, leading to slope failures [44]. Soil composition also impacts landslide susceptibility, with coarse-grained soils like gravel being more prone to failure compared to fine-grained soils such as silt and clay [45]. Geological factors further influence susceptibility, with variations in lithology, structural features, and tectonic history playing crucial roles in slope stability. Detailed geological maps that include lithological and structural characteristics are essential for accurate landslide susceptibility assessments [46].

3.1.1. Slope

Slope-derived factors provide the primary information about landslides. The diverse range of slope movements reflects the various conditions leading to slope instability and the underlying processes driving these movements [2]. Within a DEM, slope is calculated by measuring the rate of elevation change over a short distance. To analyze slope characteristics within the study area, we processed the 10 m resolution USGS 3DEP DEM in a GIS environment. Based on their degree of steepness, these slope values were classified into five distinct classes: 0–5° for low slopes, 5–20° for moderate slopes, 20–35° for steep slopes, 35–50° for very steep slopes, and $\geq 50^\circ$ for extremely steep slopes.

3.1.2. Aspect

Aspect refers to the orientation of a slope, measured clockwise from 0 to 360°, where 0° is north-facing, 90° is east-facing, 180° is south-facing, and 270° is west-facing. This classification scheme indicates how terrain features are oriented in relation to external environmental influences.

3.1.3. Curvature

The concavity or convexity of a terrain is determined by the curvature of the surface. Curvature values are calculated from the DEM. These values are then classified as follows: values less than -0.001 are classified as concave, values between -0.001 and 0.001 are classified as linear, and values greater than 0.001 are classified as convex. Concave terrain features may accumulate water and debris, increasing the likelihood of landslides, while convex terrain features facilitate surface runoff, reducing the likelihood of saturation-induced landslides.

3.1.4. Flow Direction

Flow direction is calculated from the DEM using an algorithm that determines the steepest descent from one cell to the next, representing the path of surface runoff. The flow direction values are then classified based on cardinal and intercardinal directions using the thresholds: 1, 2, 4, 8, 16, 32, 64, and 128. This classification scheme provides insights into the predominant flow patterns of surface water runoff across the study area. Analyzing the flow direction classes allows us to understand the hydrological processes shaping the landscape.

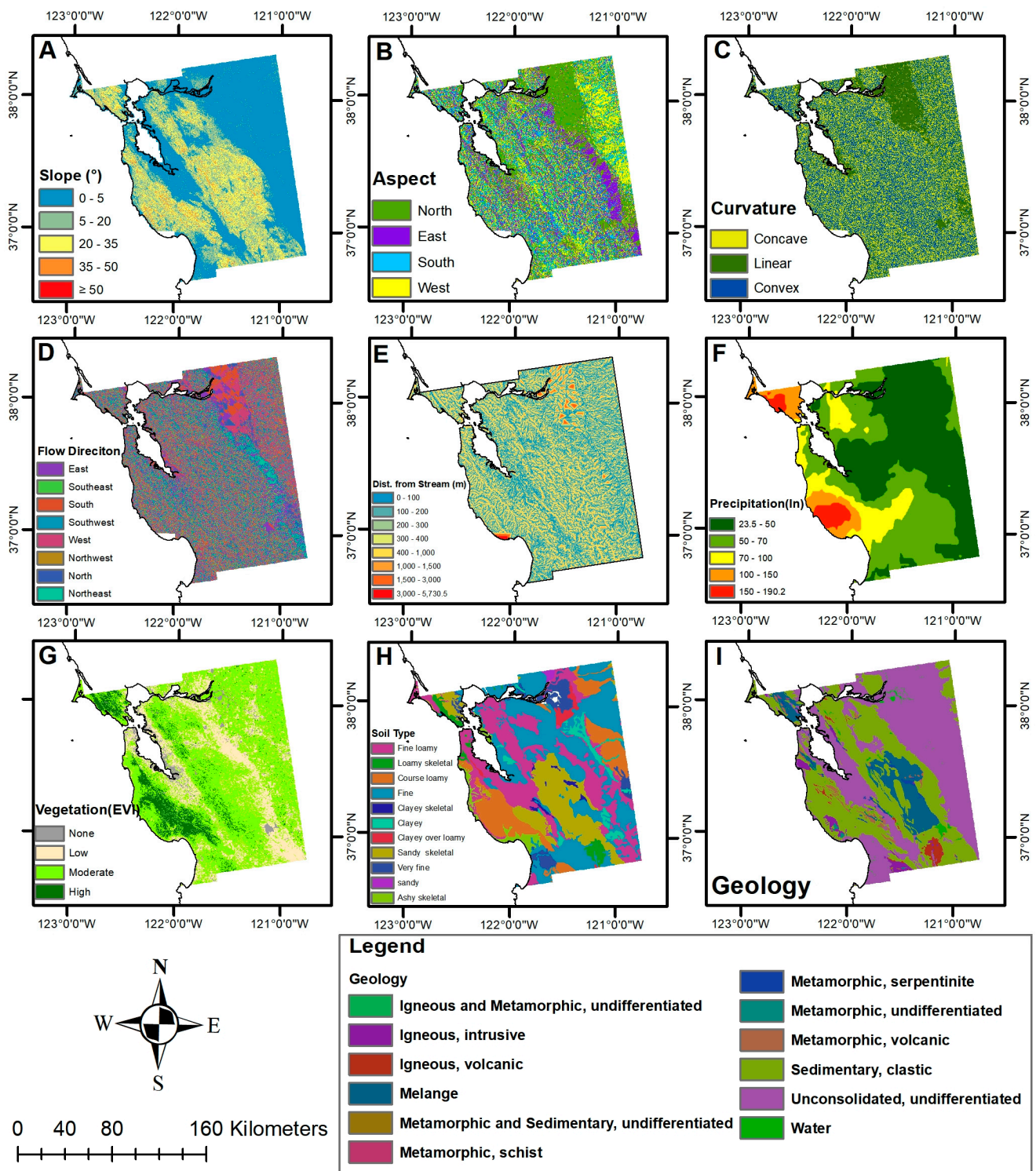


Figure 3. Different landslide causative factors used as input to the machine learning model. (A) slope, (B) aspect, (C) curvature, (D) flow direction, (E) distance from a stream, (F) rainfall, (G) vegetation, (H) soil type, and (I) geology.

3.1.5. Distance to a Stream

A catchment area is a hydrological unit determined by the direction of flow and stream order, which classifies streams based on their tributary numbers. Flow accumulation analysis identifies potential stream networks by calculating the water accumulation through each cell in the raster grid. These values are classified into two categories: from 0 to 3000

for low accumulation areas and over 3000 for high accumulation areas. A binary raster is created for catchment zones with values of 3000 or greater, and a stream network delineation algorithm generates stream order from these zones. The resulting stream network raster is converted to vector format. Finally, Euclidean distance analysis determines the distance from each cell to the nearest stream, providing insights into watercourse proximity.

3.1.6. Rainfall

Precipitation data were collected for the period 2019 to 2022 from the Community Collaborative Rain, Hail, and Snow Network (<https://www.cocorahs.org>, accessed on 13 September 2023). To obtain the spatial distribution of rainfall, a spline polynomial interpolation method was employed. This method fits a minimum curvature surface through the input data points, making it particularly suitable for spatially interpolating rainfall data. The highest recorded value was 4833.37 mm, and the lowest was 596.90 mm. To determine the overall precipitation regime within the study area, the total precipitation data for each station were averaged.

3.1.7. Vegetation

Sentinel-2 images with less than 5% cloud coverage, overlapping with the InSAR analysis period, were used for vegetation analysis. The Normalized Difference Vegetation Index (NDVI) was calculated using Google Earth Engine. NDVI values were classified into four distinct categories based on vegetation density: none, low, moderate, and high.

3.1.8. Soil Type

As soil type directly influences the hydrological and mechanical properties of a terrain, it determines landslide susceptibility. Various soil types were considered in this study, including fine loamy, loamy skeletal, coarse loamy, fine clayey skeletal, clayey over loamy, sandy skeletal, very fine, sandy, and ashy skeletal. Soil classification affects water retention capacity, permeability, and structural stability, affecting landslide susceptibility. Clayey soils, for instance, have low permeability and high water retention, which can increase pore pressure and reduce slope stability, while sandy soils have better drainage, which may help prevent landslides. Under certain conditions, loamy soils can offer moderate stability but are still susceptible to erosion and movement. Mapping these soil types across the study area can provide valuable insights into regions where soil composition may exacerbate landslides.

3.1.9. Geology

The geological profile of the study area was obtained from the USGS website [43], based on a 1:250,000 scale digitized map depicting the principal geotectonic zones, formations, clusters, and sequences. In the San Francisco area, sedimentary deposits are dominated by Franciscan/Mesozoic rock, prominently shaping the city's surface. Tertiary deposits are less common and are scattered throughout San Francisco, particularly along the coastline. Quaternary deposits are primarily found inland, especially on the middle to eastern side of the region. Surrounding San Francisco Bay, bay mud deposits are prevalent.

This region's geological map highlights various formations, each contributing to the complexity of the landscape and susceptibility to landslides. These formations include igneous and metamorphic rocks, undifferentiated metamorphic rocks, igneous intrusive and volcanic rocks, melange, schist, serpentinite, and various sedimentary formations.

3.2. MT-InSAR Data

Seventy-nine ascending passes of Sentinel-1 single look complex (SLC) images, spanning from 3 January 2019 to 31 October 2021, were utilized for MT-InSAR analysis. Our study employed a comparative approach, using the SBAS technique in one half of the study area and PSI in the other, to assess their impact on landslide susceptibility mapping. The aim was to evaluate how these MT-InSAR techniques influenced the multi-class

classification model developed for this study, focusing on their respective deformation measurements. The processing was carried out using the ENVI SARscape software v5.7.

4. Methodology

4.1. MT-InSAR

Sentinel-1 SLC images are processed using PSI and SBAS techniques. Initially, connection graphs are formed by linking each image with a reference image for the PSI technique. Interferograms are generated for each connection, and the topographic phase is compensated using a Shuttle Radar Topographic Mission (SRTM) 30 m DEM. These interferograms are then inverted to retrieve displacement rates and finally geocoded.

A similar approach is applied for SBAS, where interferograms are formed with a temporal baseline limited to 90 days. There is no limitation on the perpendicular baseline since the Sentinel-1 constellation operates within a close orbital tube. The topographic phase is compensated using a SRTM 30 m DEM. The differential interferograms are then phase-filtered to improve the signal-to-noise ratio. The Minimum Cost Flow (MCF) algorithm is employed for phase unwrapping. The unwrapped interferograms are then inverted to retrieve displacement rates. The methodology followed for landslide susceptibility mapping is illustrated in Figure 4.

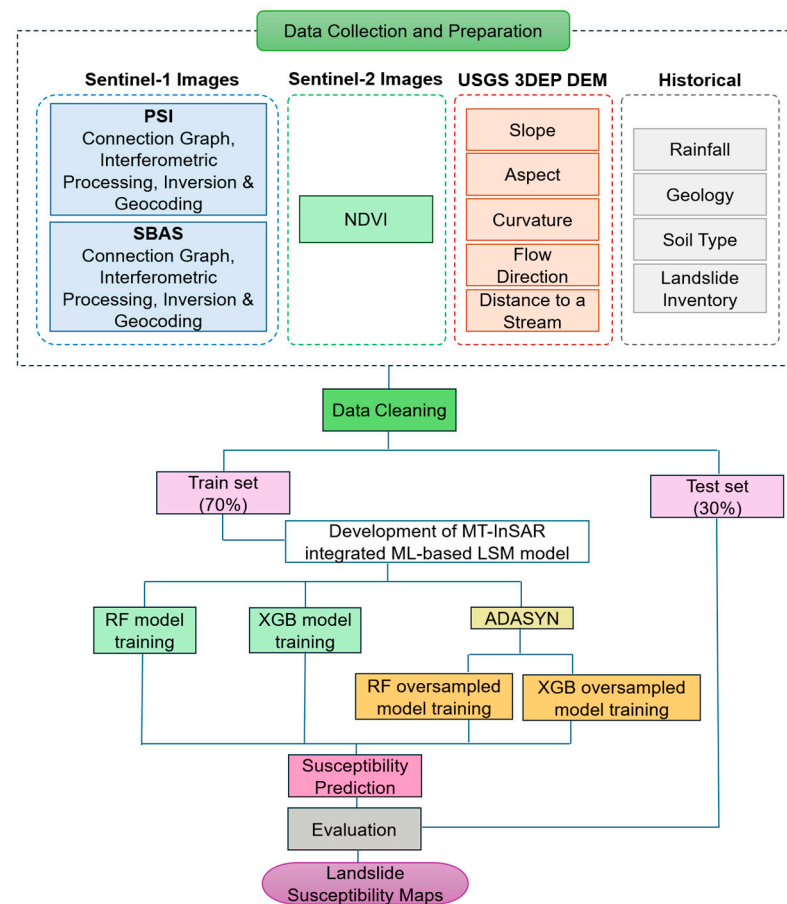


Figure 4. Flowchart of the proposed method.

4.2. Classification Criteria

The classification criteria for landslide susceptibility used in our analysis are based on rigorous scientific principles that establish a clear correlation between ground movement velocity and the likelihood of landslide occurrence [33]. Recognizing the variability in landslide occurrence across different terrains and conditions, we divided the study area into four distinct susceptibility classes. These classifications are based on the absolute velocity of

ground movement derived using MT-InSAR methods, providing a precise, velocity-based measure to gauge potential landslide activity.

The categorization ranges from Class 1, indicating areas with virtually “no susceptibility at this scale and with the available information”, to Class 4, indicating regions with “high susceptibility”. Importantly, historical landslide events, regardless of their recorded velocity, are automatically classified as Class 4. This decision is based on the premise that areas with a history of landslides, regardless of their current ground movement velocity, possess inherent geological or hydrological characteristics that predispose them to future landslide events. The classification criteria are shown in Table 1.

Table 1. Landslide susceptibility classification criteria based on absolute velocity, slope, and historical landslide events.

Class	Criteria	Susceptibility
1	Absolute velocity interval: from 0 to 2 mm/year AND Slope interval: (0, 5]°	No susceptibility at this scale, and with the available information
2	Absolute velocity interval: from 2 to 5 mm/year AND Slope interval: (5, 90]°	Low susceptibility
3	Absolute velocity interval: (5, 15] mm/year AND Slope interval: (5, 90]°	Moderate susceptibility
4	Absolute velocity interval: (>15) mm/year AND Slope interval: (5, 90]° OR Historical landslide event AND Slope interval: (5, 90]°	High susceptibility

This comprehensive and methodologically sound approach ensures that the landslide susceptibility map we have developed is informed not only by empirical data but also by a deep understanding of the landscape’s inherent susceptibility to landslides.

The thresholds were set to give the clearest picture of landslide susceptibility. Most observations fell into Class 1, showing minimal susceptibility. Class 4, indicating the highest susceptibility and including active and known landslides, made up only a small part of the dataset. This left an intermediate set of data, identifying locations that could pose a landslide threat. After the classification, Classes 2 and 3 show about the same number of data points.

A recent study [47] used the Jenks natural breaks classification method. This statistical method helps identify optimal breakpoints in data by minimizing variance within classes and maximizing it between them. This study established thresholds at 14.88 mm/year, 9.14 mm/year, 5.39 mm/year, and 2.46 mm/year, which are similar to the rounded values we determined manually, further supporting our threshold values.

Moretto [23] characterized the MT-InSAR velocities into three clusters with velocities >5 mm/yr, between 5 and 3 mm/yr, and below 3 mm/yr. The authors considered the points below 3 mm/yr as stable points. The results presented by the authors coincided with the non-linear downslope movements, linear movements, and stable areas, respectively. In addition, the authors highlight acceleration corresponding to the first cluster and forecast the time-to-failure window of the landslide. This observation will further support our classification criteria of the MT-InSAR results. These velocity thresholds are compared with the thresholds used in this study in Table 2.

Table 2. Landslide susceptibility/deformation classification criteria of this study compared with previous literature.

Study	Velocity Classification	Susceptibility
This study	0–2 mm/y	No susceptibility at this scale, and with the available information
	2–5 mm/y	Low susceptibility
	5–15 mm/y	Moderate susceptibility
	>15 mm/y	High susceptibility
Yao et al. [47]	<2.46 mm/y	Very low susceptibility
	5.39 mm/y to 2.46 mm/y	Low susceptibility
	9.14 mm/y to 5.39 mm/y	Moderate susceptibility
	14.88 mm/y to 9.14 mm/y	High susceptibility
Moretto et al. [23]	>14.88 mm/yr	Very high susceptibility
	<3 mm/y	Stable
	3–5 mm/y	Moderate deformation
	>5 mm/y	High deformation

We classify areas with a slope of 5° or less as Class 1, mainly covering relatively flat urban zones. The choice of 5° as the threshold is based on geomorphological studies suggesting that flatter terrains, typically under 5° , exhibit minimal gravitational pull that could lead to landslides. While a study using GRASS GIS for multi-scale geomorphometric analysis [48] suggests a 6° threshold to delineate relatively flat surfaces based on visual evaluations and field mapping, we opted for a slightly lower threshold to adopt a conservative approach to susceptibility mapping. This ensures the inclusion of potentially susceptible areas, enhancing our model's sensitivity to regions where landslide occurrence might be less obvious but still present.

After establishing our classification criteria based on absolute ground movement velocity and historical landslide events, our methodology includes a detailed examination of environmental and geological factors as outlined in Section 3. For each location identified through MT-InSAR data or historical records, these factors are carefully analyzed to determine their relationship with the assigned susceptibility class. This approach allows for a comprehensive correlation study, aiming to identify significant predictors of landslide susceptibility in a given area.

4.3. Ensemble Learning Models

Ensemble learning is a dynamic machine learning approach that has demonstrated clear advantages across various fields, including remote sensing. In machine learning, an ensemble refers to a system built from multiple individual models, whose outputs are combined to produce a single result for a given problem. One key benefit of using an ensemble of multiple models is that it generally offers better generalization capability compared to any single model on its own [49].

4.3.1. Random Forest (RF)

Random forests are supervised learning algorithms that employ an ensemble approach. They construct a multitude of decision trees, each trained on a bootstrap sample of the data and using a random subset of features for splitting. This random selection process reduces the correlation between trees, leading to improved generalization performance. The final prediction is determined by aggregating the predictions from all trees, often through majority voting or averaging [50]. This ensemble approach helps to mitigate the risk of overfitting and enhances the model's robustness to noise and outliers. Random forests

typically exhibit superior classification accuracy and generalization performance compared to individual decision trees [51]. One of the significant advantages of random forests is their relative insensitivity to hyperparameter tuning. The algorithm's inherent randomness and ensemble nature often yield good results with minimal parameter adjustments, making it a popular choice for various machine learning tasks.

4.3.2. Extreme Gradient Boosting (XGB)

Extreme Gradient Boosting is a supervised learning algorithm that, like decision trees, can be used for both classification and regression tasks. It is a scalable and effective version of gradient boosting [52]. The core idea behind boosting is to improve performance by combining the outputs of multiple weak learners [53], which individually have low accuracy, to create a strong classifier with enhanced performance. XGB excels in handling large datasets by building and running boosted trees in parallel, which speeds up the algorithm. One of its main advantages is its ability to optimize memory usage and leverage hardware capabilities, resulting in faster execution and better performance compared to many other machine learning algorithms [54]. Additionally, XGB incorporates L1 and L2 regularization techniques to refine the learning weights and reduce the risk of overfitting [55].

4.4. Model Parameter Tuning

Hyperparameters in machine learning algorithms vary from model to model and need to be carefully tuned before training [56]. In this study, despite the strengths of the RF and XGB models used, hyperparameter tuning was performed to enhance their performance. Specifically for RF with oversampling, several parameters were adjusted prior to training. The "class_weight" parameter was set to "balanced" rather than the default "none". Additionally, both "min_samples_leaf" and "min_samples_split" were set to 5. For XGB, the "objective" parameter was configured to "multi:softmax", and "num_class" was set to 4 to match the four-class outcome of the current problem. All other parameters were kept at their default values for each model.

4.5. Model Training

We adopted a systematic approach to model development, beginning with data partitioning. A 10% data sample was used to manage computational load without compromising the model's predictive capability. We employed ADASYN (Adaptive Synthetic Sampling) oversampling to address class imbalance, enhancing the model's ability to predict minority classes accurately. Resampling techniques, including oversampling and under-sampling, are commonly employed to address class imbalances in datasets. Under-sampling risks discarding valuable data that is crucial for model performance [56]. The ADASYN oversampling technique generates synthetic samples for the minority class based on the density distribution of existing instances, with the goal of enhancing model classification accuracy [51]. In order to evaluate the models, we used the hold-out approach, splitting the data 70% for training and 30% for testing.

To identify the most effective method for classifying landslide susceptibility, we experimented with various machine learning techniques, including RF and XGB. Our evaluation process involved experimenting with different combinations of features, using both the original dataset and an oversampled dataset to address data imbalance.

4.6. Model Validation

We used ROC (receiver operating characteristic) curves and a confusion matrix to quantitatively evaluate the accuracy and performance of the models. The model's accuracy is assessed through the AUC (area under the ROC curve). AUC values range from 0.5 to 1, with higher values closer to 1 indicating better model accuracy, while values near 0.5 suggest the model lacks predictive capability.

The confusion matrix categorizes binary classification results into four classes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). From the

parameters derived from the confusion matrix, four model performance evaluation metrics are calculated. The metrics are accuracy, precision, recall, and F1 score.

5. Results and Discussion

5.1. MT-InSAR Results

Surface deformation in the San Francisco Bay Area is primarily attributed to slow-moving geological processes such as interseismic strain accumulation, shallow fault creep, and aquifer deformation [42]. MT-InSAR results reveal a complex pattern of ground movement as shown in Figure 5. The western Santa Clara Valley is subsiding at a rate of -12 to -18 mm/year, while the eastern side exhibits a slower rate of -2 to $+12$ mm/year. A clear transition between these zones occurs northwest of the Silver Creek Fault (SCF).

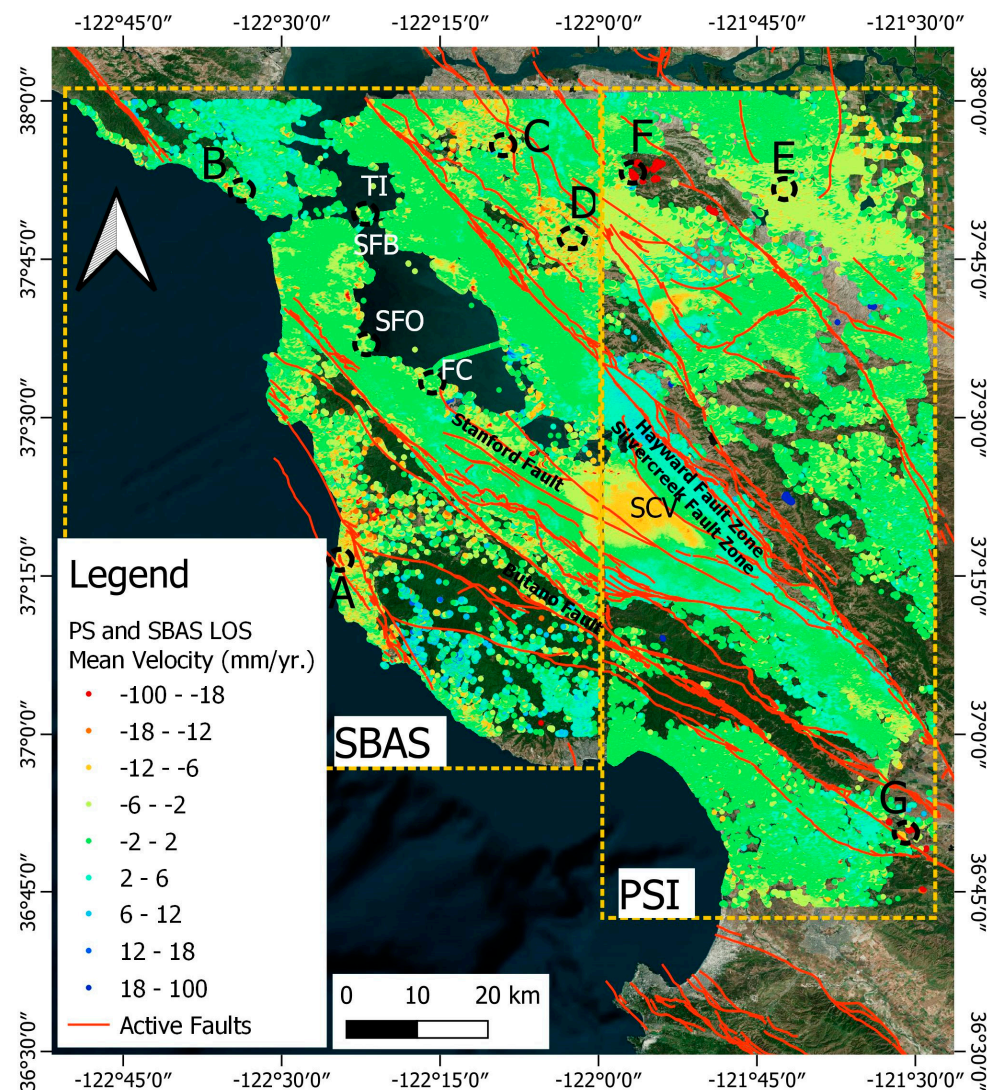


Figure 5. MT-InSAR line-of-sight mean velocity map indicating deformation rates in the San Francisco area, with Google Earth Imagery as the background. Displacement time series plots at selected locations, including Treasure Island (TI), San Francisco International Airport (SFO), Foster City (FC), and within Santa Clara Valley (SCV), are shown in Figure 6. Additional displacement time series plots at historical landslide locations (from A to G) are shown in Figure 7. The fault layer overlaid on the velocity map is retrieved from the California State Geoportal (<https://gis.data.ca.gov/>, accessed on 11 August 2023).

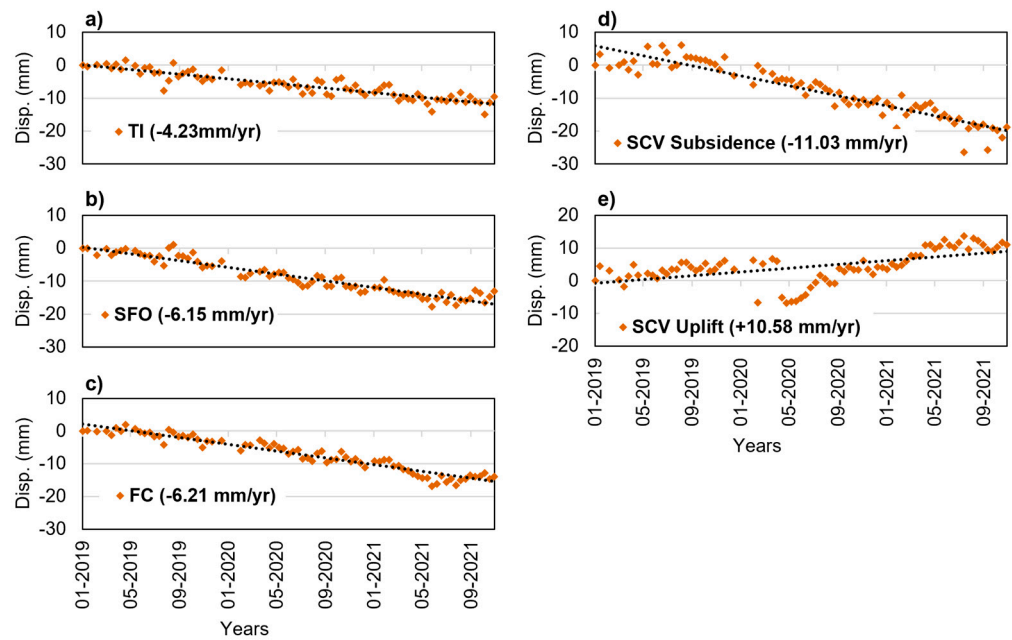


Figure 6. MT-InSAR displacement time series at (a) Treasure Island (TI), (b) San Francisco International Airport (SFO), (c) Foster City (FC), and (d,e) within Santa Clara Valley (SCV).

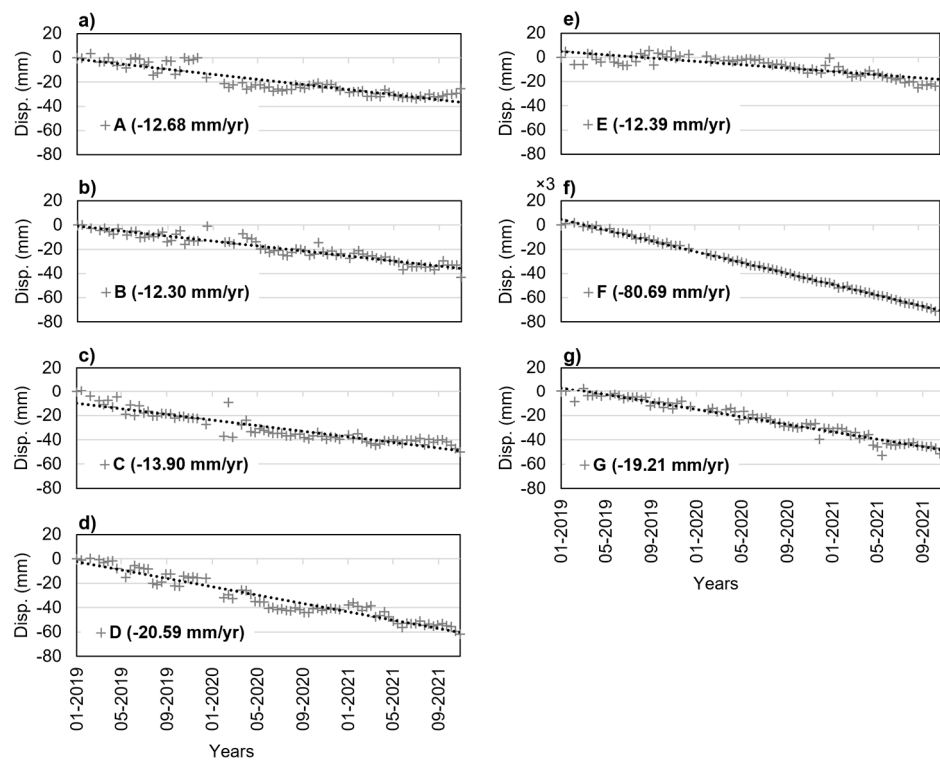


Figure 7. (a–g) MT-InSAR displacement time series at historical landslide locations denoted with letters A to G in Figure 5. The area corresponding to the high deformation rate in (f) is shown in Figure 8.

The area along the Pacific coast and around San Francisco Bay is generally stable, with deformation rates generally ranging between -2 and $+2 \text{ mm/year}$. However, specific regions, particularly those underlain by artificial landfills and mud deposits, exhibit more significant subsidence, reaching rates of -15 to -20 mm/year . Areas northwest of Treasure Island, San Francisco International Airport (SFO), and Foster City are subsiding at rates

of -2 to -6 mm/year. North of SFO, closer to the coast, subsidence rates increase to -6 to -12 mm/year. Our measured surface motion rates are consistent with previously documented deformation features [42,57,58]. Figure 6 presents displacement time series at several locations, including Treasure Island (TI), San Francisco International Airport (SFO), Foster City (FC), and within Santa Clara Valley (SCV).

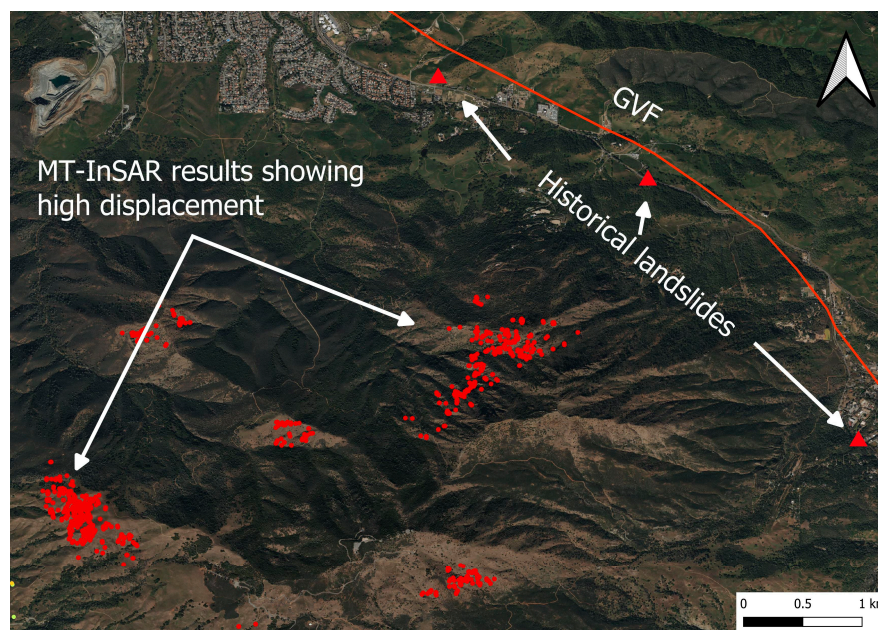


Figure 8. An area indicating a high deformation rate (subsidence) in the MT-InSAR analysis is observed near historical landslide locations (marked by red triangles) adjacent to the Green Valley Fault (GVF).

The high deformation rates observed in the MT-InSAR results may indicate areas of significant deformation on the slope that could potentially lead to landslides in the future. Figure 7 depicts deformation time series at several of these locations (represented with letters from A to G in Figure 5), showing subsidence rates ranging from -10 mm/year to -80 mm/year. An area corresponding to the high deformation is shown in Figure 8, which also highlights the presence of three historical landslides adjacent to the Green Valley Fault. Notably, this high deformation zone is located near historical landslide sites, underscoring the importance of continuous monitoring in these areas to detect and address potential landslides.

5.2. Model Accuracy Verification

The AUC value from the ROC curve provides an insightful interpretation of predictive accuracy. Figure 9 shows AUC values for each class (from Class 1 to Class 4) for the RF and XGB models with both original and oversampled data. The AUC values are between 0.92 and 0.98 in all cases, with slight variations between the original and oversampled data. The interpretations from the ROC curve suggest that both models with original and oversampled data are performing well. This result emphasizes the importance of addressing class imbalances before inputting data into machine learning models.

However, metrics such as accuracy, recall, precision, and F1 score derived from the confusion matrix reveal additional details about model performance. In addition, we also use training speed as a metric to evaluate model performance. The training speed is represented as $1/\ln(t)$, where t is time required to train the model. Models with higher accuracy, an F1 score, and a balance between recall and precision are typically regarded as good models [27]. Figure 10 portrays the accuracy, recall, precision, F1 score, and training speed for RF and XGB models. The figure clearly indicates that models with original data

do not show a balance between recall and precision, whereas models using oversampled data perform better. Both RF and XGB models with oversampled data show similar values across all four metrics. However, the training speed for the RF model with oversampled data is 16.73 min, which is much higher than the training speed for the XGB model with oversampled dataset (1.54 min). Even though the training speed for the XGB model with original data (0.53 min) is lower than XGB with oversampled data, the balance between recall and precision values is not promising. Therefore, the XGB model with oversampled data for landslide susceptibility mapping has been selected in this study as it demonstrated better performance.

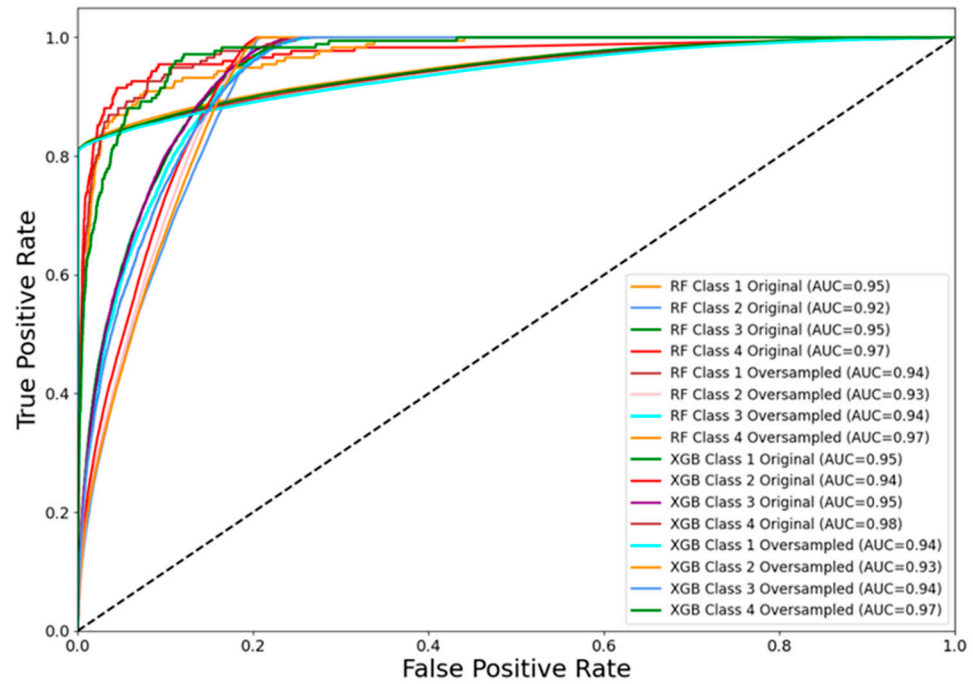


Figure 9. ROC curve and AUC values for each class (from Class 1 to Class 4) using RF and XGB models with original and oversampled data.

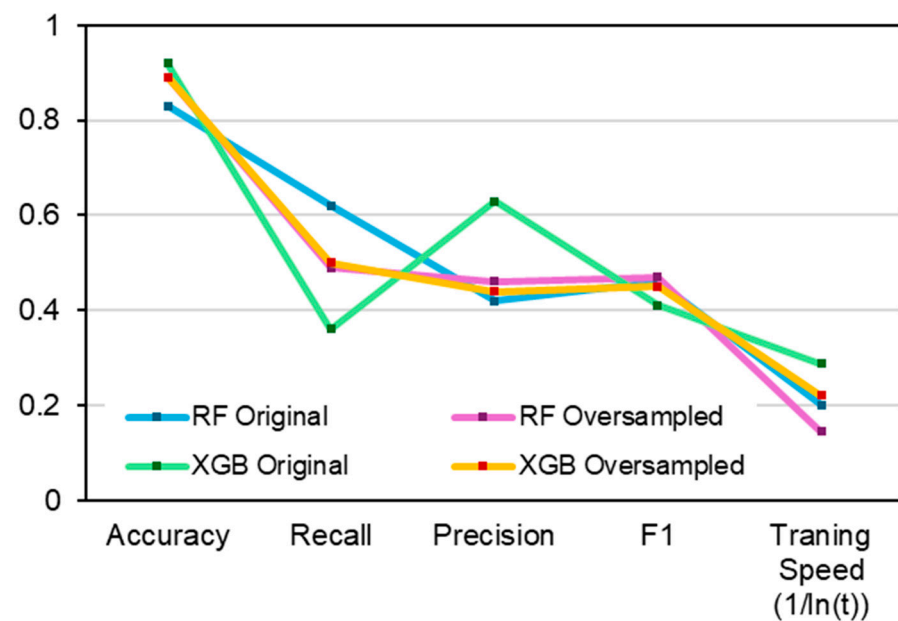


Figure 10. Comparative metrics for RF and XGB models with original and oversampled data.

5.3. Landslide Susceptibility Mapping

Landslide susceptibility maps identify areas that are prone to landslides, categorizing them by susceptibility level. These maps are invaluable for government and private agencies in land-use planning, infrastructure development, and disaster management. By pinpointing high-susceptible zones, authorities can implement preventive measures such as reforestation, drainage systems, and early warning systems. During disasters, these maps aid in rescue efforts, resource allocation, and evacuation planning, minimizing loss of life and property damage.

Factors such as slope, vegetation, and proximity to streams contribute to landslide susceptibility, collectively accounting for 79%. A tree map showing the importance of factors used in this study is shown in Figure 11. Among these, slope is the primary factor, accounting for 37% of landslide susceptibility. Slope has been identified as a crucial factor in landslide susceptibility in numerous mapping studies [2,9–11], particularly affecting susceptibility at steeper angles. Vegetation accounts for approximately 22% of landslide susceptibility. Low to moderate susceptibility is observed in areas with high vegetation cover, indicated by high NDVI values. Proximity to streams contributes to 20% of landslide susceptibility, as these areas often experience soil saturation, frequent flooding, and soil erosion, which increase the risk of landslides. Geology, rainfall, and soil type together contribute about 16%. Variations in rainfall intensity and patterns can notably affect the timing and characteristics of landslides [59,60]. For example, a time lag between rainfall events and landslides has been observed in the Berkeley Hills of California [35,36,41]. It was observed that both low and high precipitation events equally contributed to landslide susceptibility in the study area. Loamy and clayey soils were found to have a greater impact on susceptibility in this study. Landslide susceptibility is primarily distributed in areas with sedimentary, clastic geology. In contrast, factors such as aspect, flow direction, and curvature had minimal influence.

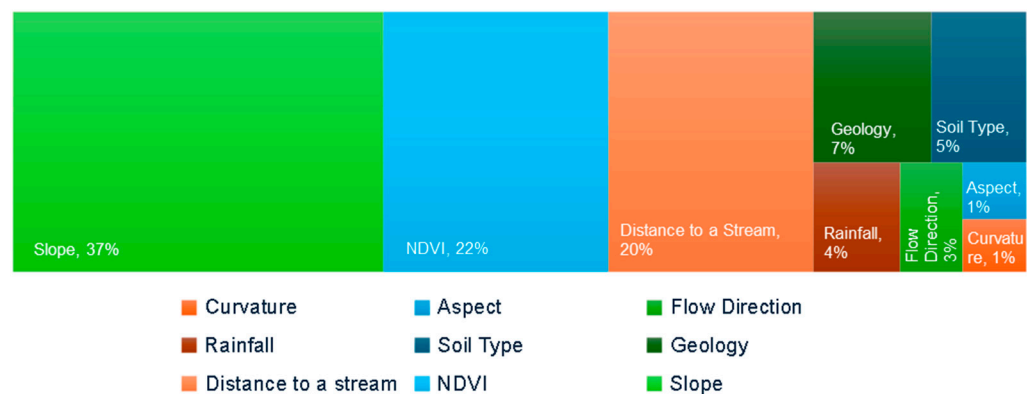


Figure 11. Tree map showing the importance of factors used in this study.

The LSM is classified into four classes ranging from Class 1 to Class 4, denoting no susceptibility at this scale and, with the available information, high susceptibility. The landslide susceptibility map generated by the XGB model is shown in Figure 12 indicates high susceptibility along the Pacific coast. The SCV area is predicted as Class 1, as the slope values are between 0° and 5° . The map's predictions align well with historical landslide locations, demonstrating the model's effectiveness in predicting landslides. The optimal model was validated against 272 historical landslide occurrences in the area of interest, with predictions distributed as follows: 68 occurrences (25%) in Class 1, 142 occurrences (52%) in Class 2, 58 occurrences (21.5%) in Class 3, and 4 occurrences (1.5%) in Class 4.

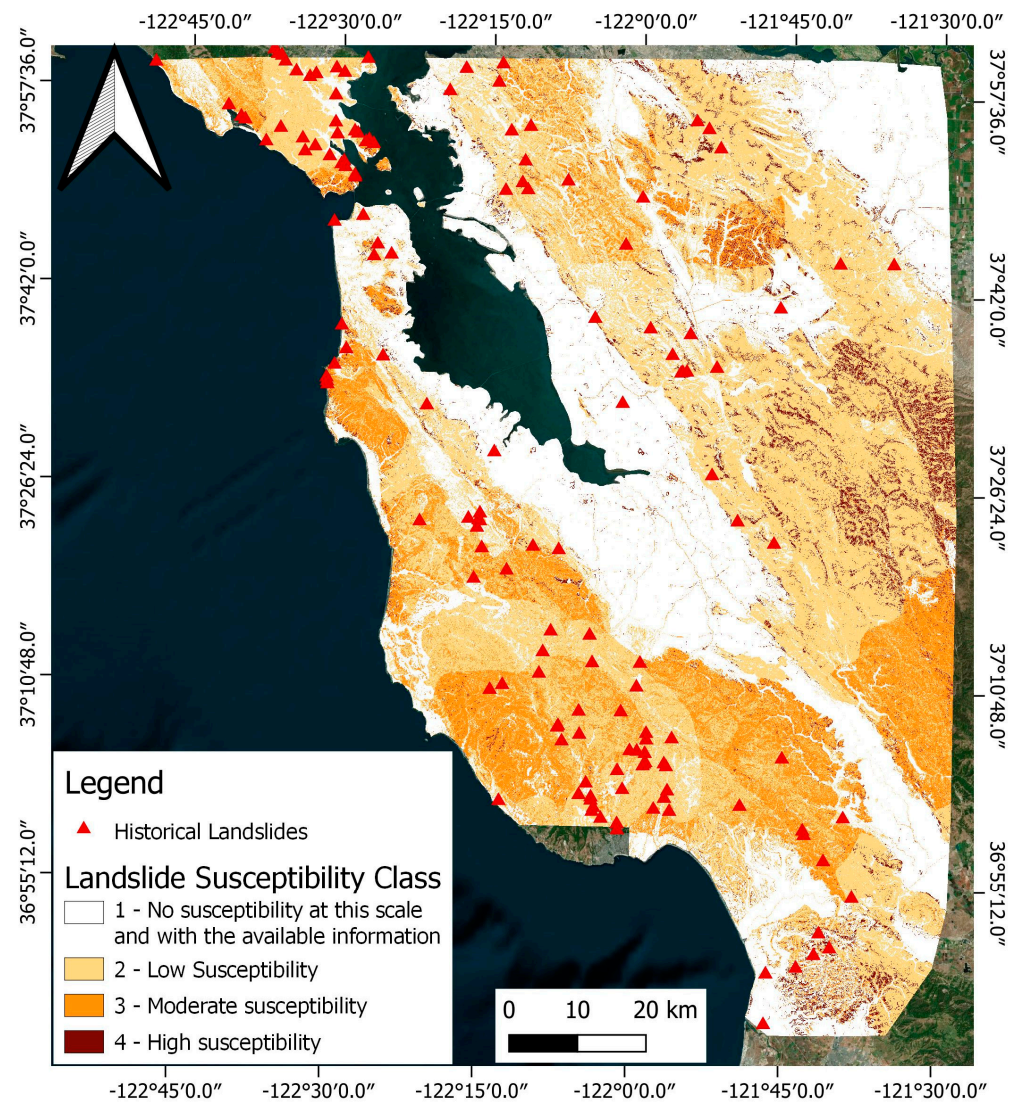


Figure 12. Landslide susceptibility map of the San Francisco area derived using the XGB model. The map classifies landslides into four categories, ranging from no susceptibility to high susceptibility. The historical landslide locations are overlaid on the susceptibility map.

This outcome highlights a crucial aspect of landslide prediction models—the dynamic nature of environmental conditions. It is vital to recognize that the factors leading to past landslide events may have significantly changed, making areas previously prone to landslides more stable today. Vegetation growth, changes in land use, and erosion control measures can greatly alter the landscape, enhancing stability and reducing landslide susceptibility. Therefore, areas that experienced landslides in the past may not necessarily present the same level of susceptibility today. Additionally, the historical accuracy of landslide occurrences might vary, with some events potentially misclassified or affected by data collection limitations at the time.

This context provides a logical framework for understanding the validation results. While the optimal model's performance against historical data might initially suggest limitations, it also underscores the importance of considering temporal changes in environmental conditions such as precipitation, distance to streams, changes in vegetation, and the evolving nature of landslide behavior. This perspective encourages continuous model refinement and highlights the need for integrating up-to-date environmental data to enhance predictive accuracy. Future iterations of the model will benefit from incorporating

recent changes in the parameters that influence landslide susceptibility, thereby improving its relevance and reliability in current and future assessments.

6. Conclusions

This study successfully developed and evaluated a landslide susceptibility model for the San Francisco Bay Area. The model integrates conventional Earth Observation (EO) data derivatives (slope, aspect, NDVI, etc.), historical geospatial data (rainfall, geology, etc.), and MT-InSAR ground movement results. It leverages the strengths of these diverse data sources to create a multi-class classification model, categorizing geographical areas into four distinct susceptibility levels: no susceptibility at this scale and with the available information, low susceptibility, medium susceptibility, and high susceptibility.

MT-InSAR analysis revealed complex patterns of ground movement in the San Francisco Bay Area, with subsidence rates ranging from -20 mm/year to $+12$ mm/year. These findings align well with existing literature. Importantly, MT-InSAR identified subsidence at historical landslide locations, highlighting its effectiveness in landslide mapping.

Machine learning models, particularly the XGB model with oversampled data, demonstrated promising performance in landslide susceptibility mapping. The model achieved high AUC values (≥ 0.93) across all susceptibility classes, indicating good predictive accuracy. Additionally, the confusion matrix metrics (accuracy, precision, recall, and F1 score) showed a balanced performance for the XGB model with oversampled data, making it the optimal model for this study. The training and predictive speed of the XGB model using oversampled data are ten times higher compared to the RF model using the same data.

The generated landslide susceptibility map classified susceptibility levels across the San Francisco Bay Area. The map identified high-susceptibility zones along the Pacific coast, which aligns well with the distribution of historical landslides. This demonstrates the model's ability to predict potential landslide occurrences. The validation process using historical landslide data revealed that a significant portion (52%) of past landslides fell into Class 2 (low susceptibility) of the model's classification. While this might initially appear as a limitation, it can be attributed to the dynamic nature of the environment. Factors like vegetation growth, land-use changes, and erosion control measures can alter the landscape over time, potentially reducing susceptibility in areas previously susceptible to landslides. Additionally, limitations in historical data accuracy might contribute to discrepancies.

Future research should explore incorporating additional dynamic factors, such as real-time precipitation data and seismic activity, into the model. This could enhance the model's ability to capture the temporal variability of landslide susceptibility. Regularly updating the model with the latest EO data and historical landslide occurrences will be crucial for maintaining its accuracy and effectiveness over time. Overall, this study presents a significant contribution to landslide susceptibility mapping by integrating MT-InSAR data and machine learning techniques. The developed model provides valuable insights for land-use planning and hazard mitigation strategies in the San Francisco Bay Area.

Author Contributions: Conceptualization, D.S.V. and S.P.; data curation, V.R.Y.; formal analysis, V.R.Y.; funding acquisition, S.P.; methodology, D.S.V., V.R.Y., S.P., and A.P.; project administration, S.P.; resources, S.P.; software, A.P.; supervision, S.P.; validation, D.S.V. and V.R.Y.; visualization, D.S.V. and V.R.Y.; writing—original draft, D.S.V. and V.R.Y.; writing—review and editing, D.S.V., V.R.Y., and S.P. All authors have read and agreed to the published version of the manuscript.

Funding: The Project INNOVATE/1221/0019 was funded by the European Union Recovery and Resilience Facility of the NextGenerationEU instrument through the Research and Innovation Foundation.

Data Availability Statement: The data are available upon request from the corresponding author.

Conflicts of Interest: The authors were employed by the company GEOFEM and declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Hussain, S.; Pan, B.; Afzal, Z.; Ali, M.; Zhang, X.; Shi, X.; Ali, M. Landslide Detection and Inventory Updating Using the Time-Series InSAR Approach along the Karakoram Highway, Northern Pakistan. *Sci. Rep.* **2023**, *13*, 7485. [[CrossRef](#)] [[PubMed](#)]
2. Popescu, M. Landslide Causal Factors and Landslide Remediation Options. In Proceedings of the 3rd International Conference on Landslides, Slope Stability and Safety of Infrastructures, Singapore, 11–12 July 2002; pp. 61–81.
3. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.-X.; Li, S. Novel Hybrid Artificial Intelligence Approach of Bivariate Statistical-Methods-Based Kernel Logistic Regression Classifier for Landslide Susceptibility Modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4397–4419. [[CrossRef](#)]
4. Nhu, V.-H.; Mohammadi, A.; Shahabi, H.; Ahmad, B.B.; Al-Ansari, N.; Shirzadi, A.; Clague, J.J.; Jaafari, A.; Chen, W.; Nguyen, H. Landslide Susceptibility Mapping Using Machine Learning Algorithms and Remote Sensing Data in a Tropical Environment. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4933. [[CrossRef](#)] [[PubMed](#)]
5. Kouhartsiouk, D.; Perdikou, S. The Application of DInSAR and Bayesian Statistics for the Assessment of Landslide Susceptibility. *Nat. Hazards* **2021**, *105*, 2957–2985. [[CrossRef](#)]
6. Brabb, E.E. The World Landslide Problem. *Epis. J. Int. Geosci.* **1991**, *14*, 52–61. [[CrossRef](#)]
7. Ghorbani, Z.; Khosravi, A.; Maghsoudi, Y.; Mojtahedi, F.F.; Javadnia, E.; Nazari, A. Use of InSAR Data for Measuring Land Subsidence Induced by Groundwater Withdrawal and Climate Change in Ardabil Plain, Iran. *Sci. Rep.* **2022**, *12*, 13998. [[CrossRef](#)]
8. Kirschbaum, D.B.; Adler, R.; Hong, Y.; Lerner-Lam, A. Evaluation of a Preliminary Satellite-Based Landslide Hazard Algorithm Using Global Landslide Inventories. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 673–686. [[CrossRef](#)]
9. Lee, S.; Min, K. Statistical Analysis of Landslide Susceptibility at Yongin, Korea. *Environ. Geol.* **2001**, *40*, 1095–1113. [[CrossRef](#)]
10. Dai, F.C.; Lee, C.F. Landslide Characteristics and Slope Instability Modeling Using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [[CrossRef](#)]
11. Coe, J.; Michael, J.; Crovelli, R.; Savage, W.; Laprade, W.; Nashem, W. Probabilistic Assessment of Precipitation-Triggered Landslides Using Historical Records of Landslide Occurrence, Seattle, Washington. *Environ. Eng. Geosci.* **2004**, *10*, 103–122. [[CrossRef](#)]
12. Ruff, M.; Czurda, K. Landslide Susceptibility Analysis with a Heuristic Approach in the Eastern Alps (Vorarlberg, Austria). *Geomorphology* **2008**, *94*, 314–324. [[CrossRef](#)]
13. Evans, H.; Pennington, C.; Jordan, C.; Foster, C. Mapping a Nation's Landslides: A Novel Multi-Stage Methodology. In *Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning*; Margottini, C., Canuti, P., Sassa, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 21–27. ISBN 978-3-642-31325-7.
14. Devaraj, S.; Yarrakula, K.; Martha, T.R.; Murugesan, G.P.; Vaka, D.S.; Surampudi, S.; Wadhwa, A.; Loganathan, P.; Budamala, V. Time Series SAR Interferometry Approach for Landslide Identification in Mountainous Areas of Western Ghats, India. *J. Earth Syst. Sci.* **2022**, *131*, 133. [[CrossRef](#)]
15. Famiglietti, N.A.; Miele, P.; Defilippi, M.; Cantone, A.; Riccardi, P.; Tessari, G.; Vicari, A. Landslide Mapping in Calitri (Southern Italy) Using New Multi-Temporal InSAR Algorithms Based on Permanent and Distributed Scatterers. *Remote Sens.* **2024**, *16*, 1610. [[CrossRef](#)]
16. Kim, J.-W.; Lu, Z.; Kaufmann, J. Evolution of Sinkholes over Wink, Texas, Observed by High-Resolution Optical and SAR Imagery. *Remote Sens. Environ.* **2019**, *222*, 119–132. [[CrossRef](#)]
17. Vaka, D.S.; Rao, Y.S.; Bhattacharya, A. Surface Displacements of the 12 November 2017 Iran–Iraq Earthquake Derived Using SAR Interferometry. *Geocarto Int.* **2021**, *36*, 660–675. [[CrossRef](#)]
18. Vaka, D.S.; Rao, Y.S.; Bhattacharya, A. Time Series Analysis of the Pre-Seismic and Post-Seismic Surface Deformation of the 2017 Iran–Iraq Earthquake Derived from Sentinel-1 InSAR Data. *J. Earth Syst. Sci.* **2023**, *132*, 64. [[CrossRef](#)]
19. Ferretti, A.; Prati, C.; Rocca, F. Permanent Scatterers in SAR Interferometry. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 8–20. [[CrossRef](#)]
20. Berardino, P.; Fornaro, G.; Lanari, R.; Sansosti, E. A New Algorithm for Surface Deformation Monitoring Based on Small Baseline Differential SAR Interferograms. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2375–2383. [[CrossRef](#)]
21. Lanari, R.; Casu, F.; Manzo, M.; Zeni, G.; Berardino, P.; Manunta, M.; Pepe, A. An Overview of the Small Baseline Subset Algorithm: A DInSAR Technique for Surface Deformation Analysis. *Pure Appl. Geophys.* **2007**, *164*, 637–661. [[CrossRef](#)]
22. Bouali, E.H.; Oommen, T.; Escobar-Wolf, R. Mapping of Slow Landslides on the Palos Verdes Peninsula Using the California Landslide Inventory and Persistent Scatterer Interferometry. *Landslides* **2018**, *15*, 439–452. [[CrossRef](#)]
23. Moretto, S.; Bozzano, F.; Mazzanti, P. The Role of Satellite InSAR for Landslide Forecasting: Limitations and Openings. *Remote Sens.* **2021**, *13*, 3735. [[CrossRef](#)]
24. Zhang, L.; Dai, K.; Deng, J.; Ge, D.; Liang, R.; Li, W.; Xu, Q. Identifying Potential Landslides by Stacking-InSAR in Southwestern China and Its Performance Comparison with SBAS-InSAR. *Remote Sens.* **2021**, *13*, 3662. [[CrossRef](#)]
25. Hussain, M.A.; Chen, Z.; Zheng, Y.; Shoaib, M.; Shah, S.U.; Ali, N.; Afzal, Z. Landslide Susceptibility Mapping Using Machine Learning Algorithm Validated by Persistent Scatterer In-SAR Technique. *Sensors* **2022**, *22*, 3119. [[CrossRef](#)] [[PubMed](#)]
26. Yao, J.; Yao, X.; Liu, X. Landslide Detection and Mapping Based on SBAS-InSAR and PS-InSAR: A Case Study in Gongjue County, Tibet, China. *Remote Sens.* **2022**, *14*, 4728. [[CrossRef](#)]
27. Miao, F.; Ruan, Q.; Wu, Y.; Qian, Z.; Kong, Z.; Qin, Z. Landslide Dynamic Susceptibility Mapping Base on Machine Learning and the PS-InSAR Coupling Model. *Remote Sens.* **2023**, *15*, 5427. [[CrossRef](#)]

28. Wu, X.; Qi, X.; Peng, B.; Wang, J. Optimized Landslide Susceptibility Mapping and Modelling Using the SBAS-InSAR Coupling Model. *Remote Sens.* **2024**, *16*, 2873. [[CrossRef](#)]
29. Sarkar, S.; Kanungo, D.P. An Integrated Approach for Landslide Susceptibility Mapping Using Remote Sensing and GIS. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 617–625. [[CrossRef](#)]
30. Tyoda, Z. Landslide Susceptibility Mapping: Remote Sensing and GIS Approach. Ph.D. Thesis, Stellenbosch University, Stellenbosch, South Africa, 2013.
31. Nhu, V.-H.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Jaafari, A.; Chen, W.; Miraki, S.; Dou, J. Shallow Landslide Susceptibility Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naïve Bayes Tree, Artificial Neural Network, and Support Vector Machine Algorithms. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2749. [[CrossRef](#)]
32. Ali, N.; Chen, J.; Fu, X.; Ali, R.; Hussain, M.A.; Daud, H.; Hussain, J.; Altalbe, A. Integrating Machine Learning Ensembles for Landslide Susceptibility Mapping in Northern Pakistan. *Remote Sens.* **2024**, *16*, 988. [[CrossRef](#)]
33. Whiteley, J.S.; Watlet, A.; Kendall, J.M.; Chambers, J.E. Brief Communication: The Role of Geophysical Imaging in Local Landslide Early Warning Systems. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 3863–3871. [[CrossRef](#)]
34. Handwerker, A.L.; Huang, M.-H.; Fielding, E.J.; Booth, A.M.; Bürgmann, R. A Shift from Drought to Extreme Rainfall Drives a Stable Landslide to Catastrophic Failure. *Sci. Rep.* **2019**, *9*, 1569. [[CrossRef](#)] [[PubMed](#)]
35. Kang, Y.; Lu, Z.; Zhao, C.; Xu, Y.; Kim, J.; Gallegos, A.J. InSAR Monitoring of Creeping Landslides in Mountainous Regions: A Case Study in Eldorado National Forest, California. *Remote Sens. Environ.* **2021**, *258*, 112400. [[CrossRef](#)]
36. Cohen-Waeber, J.; Sitar, N.; Bürgmann, R. GPS Instrumentation and Remote Sensing Study of Slow Moving Landslides in the Eastern San Francisco Bay Hills, California, USA. In Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris, France, 2–6 September 2013; pp. 2169–2172.
37. Cohen-Waeber, J.; Bürgmann, R.; Sitar, N.; Ferretti, A.; Giannico, C.; Bianchi, M. *18 Geodetic Tracking and Characterization of Precipitation Triggered Slow Moving Landslide Displacements in the Eastern San Francisco Bay Hills, California, USA*; Berkeley Seismological Laboratory: Berkeley, CA, USA, 2013; Volume 42.
38. Hilley, G.E.; Bürgmann, R.; Ferretti, A.; Novali, F.; Rocca, F. Dynamics of Slow-Moving Landslides from Permanent Scatterer Analysis. *Science* **2004**, *304*, 1952–1955. [[CrossRef](#)] [[PubMed](#)]
39. Quigley, K.C.; Bürgmann, R.; Giannico, C.; Novali, F.; Knudsen, I.K. Seasonal Acceleration and Structure of Slow Moving Landslides in the Berkeley Hills. In *California Geological Survey Special Report, Proceedings of the Third Conference on Earthquake Hazards, Eastern San Francisco Bay Area, CA, USA, 22–24 October 2008*; California Department of Conservation: Sacramento, CA, USA, 2010; Volume 219.
40. Giannico, C.; Ferretti, A. *SqueeSAR Analysis Area: Berkeley*; Tele-Rilevamento Europa: Milano, Italy, 2011.
41. Cohen-Waeber, J.; Bürgmann, R.; Chaussard, E.; Giannico, C.; Ferretti, A. Spatiotemporal Patterns of Precipitation-Modulated Landslide Deformation from Independent Component Analysis of InSAR Time Series. *Geophys. Res. Lett.* **2018**, *45*, 1878–1887. [[CrossRef](#)]
42. Bürgmann, R.; Hilley, G.; Ferretti, A.; Novali, F. Resolving Vertical Tectonics in the San Francisco Bay Area from Permanent Scatterer InSAR and GPS Analysis. *Geology* **2006**, *34*, 221–224. [[CrossRef](#)]
43. Horton, J.D.; San Juan, C.A.; Stoesser, D.B. *The State Geologic Map Compilation (SGMC) Geodatabase of the Conterminous United States (Version 1.1, August 2017)*; U.S. Geological Survey Data Series; U.S. Geological Survey: Washington, DC, USA, 2017; Volume 1052, 46p. [[CrossRef](#)]
44. Huang, G.; Zheng, M.; Peng, J. Effect of Vegetation Roots on the Threshold of Slope Instability Induced by Rainfall and Runoff. *Geofluids* **2021**, *2021*, 6682113. [[CrossRef](#)]
45. Liu, Y.; Deng, Z.; Wang, X. The Effects of Rainfall, Soil Type and Slope on the Processes and Mechanisms of Rainfall-Induced Shallow Landslides. *Appl. Sci.* **2021**, *11*, 11652. [[CrossRef](#)]
46. Segoni, S.; Pappafico, G.; Luti, T.; Catani, F. Landslide Susceptibility Assessment in Complex Geological Settings: Sensitivity to Geological Information and Insights on Its Parameterization. *Landslides* **2020**, *17*, 2443–2453. [[CrossRef](#)]
47. Yao, Z.; Chen, M.; Zhan, J.; Zhuang, J.; Sun, Y.; Yu, Q.; Yu, Z. Refined Landslide Susceptibility Mapping by Integrating the SHAP-CatBoost Model and InSAR Observations: A Case Study of Lishui, Southern China. *Appl. Sci.* **2023**, *13*, 12817. [[CrossRef](#)]
48. Veselský, M.; Bandura, P.; Burian, L.; Hrciníková, T.; Bella, P. Semi-Automated Recognition of Planation Surfaces and Other Flat Landforms: A Case Study from the Aggtelek Karst, Hungary. *Open Geosci.* **2015**, *7*, 63. [[CrossRef](#)]
49. Huang, F.; Xie, G.; Xiao, R. Research on Ensemble Learning. In Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence, Shanghai, China, 7–8 November 2009; Volume 3, pp. 249–252.
50. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
51. Nam, K.; Kim, J.; Chae, B.-G. Exploring Class Imbalance with Under-Sampling, over-Sampling, and Hybrid Sampling Based on Mahalanobis Distance for Landslide Susceptibility Assessment: A Case Study of the 2018 Iburi Earthquake Induced Landslides in Hokkaido, Japan. *Geosci. J.* **2024**, *28*, 71–94. [[CrossRef](#)]
52. Zhang, Y.; Liang, S.; Zhu, Z.; Ma, H.; He, T. Soil Moisture Content Retrieval from Landsat 8 Data Using Ensemble Learning. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 32–47. [[CrossRef](#)]
53. Zhang, W.; He, Y.; Wang, L.; Liu, S.; Meng, X. Landslide Susceptibility Mapping Using Random Forest and Extreme Gradient Boosting: A Case Study of Fengjie, Chongqing. *Geol. J.* **2023**, *58*, 2372–2387. [[CrossRef](#)]

54. Dhieb, N.; Ghazzai, H.; Besbes, H.; Massoud, Y. Extreme Gradient Boosting Machine Learning Algorithm for Safe Auto Insurance Operations. In Proceedings of the 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES), Cairo, Egypt, 4–6 September 2019; pp. 1–5.
55. Ahmad, M.; Al-Mansob, R.A.; Kashyzadeh, K.R.; Keawsawasvong, S.; Sabri Sabri, M.M.; Jamil, I.; Alguno, A.C. Extreme Gradient Boosting Algorithm for Predicting Shear Strengths of Rockfill Materials. *Complexity* **2022**, *2022*, 9415863. [[CrossRef](#)]
56. Sharma, N.; Saharia, M.; Ramana, G.V. High Resolution Landslide Susceptibility Mapping Using Ensemble Machine Learning and Geospatial Big Data. *Catena* **2024**, *235*, 107653. [[CrossRef](#)]
57. Shirzaei, M.; Bürgmann, R.; Fielding, E.J. Applicability of Sentinel-1 Terrain Observation by Progressive Scans Multitemporal Interferometry for Monitoring Slow Ground Motions in the San Francisco Bay Area. *Geophys. Res. Lett.* **2017**, *44*, 2733–2742. [[CrossRef](#)]
58. Shirzaei, M.; Bürgmann, R. Global Climate Change and Local Land Subsidence Exacerbate Inundation Risk to the San Francisco Bay Area. *Sci. Adv.* **2018**, *4*, eaap9234. [[CrossRef](#)]
59. Li, Q.; Huang, D.; Pei, S.; Qiao, J.; Wang, M. Using Physical Model Experiments for Hazards Assessment of Rainfall-Induced Debris Landslides. *J. Earth Sci.* **2021**, *32*, 1113–1128. [[CrossRef](#)]
60. Wu, L.Z.; Xu, Q.; Zhu, J.D. Incorporating Hydro-Mechanical Coupling in an Analysis of the Effects of Rainfall Patterns on Unsaturated Soil Slope Stability. *Arab. J. Geosci.* **2017**, *10*, 386. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.