Preprints.org

Article

Machine Learning Analysis of Precipitation as Trigger for Shallow Landslides: Ometepe Island, Nicaragua

Victor Arcia Castro * and Gerald Corzo

Article

Machine Learning Analysis of Precipitation as Trigger for Shallow Landslides: Ometepe Island, Nicaragua

Victor Arcia *,†,‡ [ID] **and Gerald Corzo** ‡ [ID]

IHE Delft
* Correspondence: victorarciac@gmail.com; Tel.: +31 (0)152151715
† Current address: Westvest 7, 2611 AX Delft
‡ These authors contributed equally to this work.

Abstract: In this study, we address the need for improved landslide forecasting by exploring innovative machine learning approaches to classify shallow landslide events. Focusing on Ometepe Island, Nicaragua, known for its susceptibility to hydrometeorological incidents and seismic activities, we constructed a comprehensive shallow landslide database using coherence layers derived from SAR images. We applied coherence thresholds ranging from 0.9 to 0.99, in conjunction with 7-day aggregated spatio-temporal precipitation data. By employing machine learning models such as Logistic Regression, Random Forest, and Support Vector Classifier, we established a correlation between precipitation patterns and landslide occurrences. The Support Vector Classifier with a Sigmoidal kernel proved to be the most effective, achieving an F1 score of 0.67. Our results demonstrate a linkage between precipitation and landslide events, confirming the critical role of rainfall over 7-day periods in triggering these events. The study also indicates that, while rainfall is a crucial factor, it is not the sole trigger for landslides, as seismic activities also play a significant role. This research contributes valuable insights into landslide prediction and risk assessment, underscoring the potential of machine learning in enhancing environmental hazard analysis.

Keywords: data-driven modelling; machine learning; landslide; remote sensing; Logistic Regression; Random Forest; Support Vector Classifier; Ometepe; Nicaragua

1. Introduction

Landslides, particularly shallow landslides, pose significant hazards worldwide, claiming more than 55,000 lives between 2004 and 2016 [1] and inflicting economic damages upwards of 4.7 billion euros annually in Europe [2]. Initiatives such as the Sendai agreement for risk reduction and the Kyoto 2020 commitment exemplify the global interest in safeguarding civilians from these catastrophes [3]. In addition to their direct impact on human life, shallow landslides can trigger environmental repercussions such as accelerated deforestation and soil fertility decline, adding further urgency to the issue [4]. Central American countries, including Nicaragua, are particularly prone to these hazards.

Predicting landslides has been a formidable challenge, primarily due to the intricate interplay of influencing factors such as precipitation, soil moisture, geological conditions, vegetation cover, human activity, and regional seismic risks. Landslides, and especially shallow ones, are difficult to detect both temporally and spatially, which significantly compounds the complexity of prediction. Additionally, the triggers are notably heterogeneous in space and occasionally in time, mandating a fusion of remote sensing data with in-situ variables.

Effective landslide prediction demands a multi-scale approach; both broad-scale and detailed analyses are equally crucial for achieving accurate forecasts [5]. However, the utilization of advanced, local information-inclusive techniques has been hindered by data limitations, especially at necessary scales and in real-time [6]. Shallow landslides introduce additional challenges due to their rapid onset and the variable nature of their triggers.

In the context of Nicaragua, the country's low per-capita income, coupled with the highest poverty rate in Central America [7], and its classification as the fourth most vulnerable country to climate change [8], exacerbate the threat of landslides. The population's vulnerability is heightened by the region's natural conditions and a lack of comprehensive databases detailing landslide events. Shallow

landslides, in particular, have caused 74.6% of natural disaster-related deaths in the last 15 years, necessitating a robust investigation into their triggers [9].

Remote sensing, coupled with machine learning techniques, has emerged as a powerful tool for landslide detection and forecasting. The integration of remote sensing data with machine learning algorithms has demonstrated promising results in enhancing the accuracy and efficiency of landslide detection [10–13].

This has prompted researchers to develop forecasting tools for risk assessment [14]. The use of multi-source remote sensing techniques and machine learning has attracted increasing attention, indicating a growing interest in leveraging diverse data sources for landslide prediction [10]. Moreover, the application of innovative machine learning algorithms, such as Logistic Regression, Random Forest and Support Vector Classifier has the potential to enhance landslide susceptibility mapping efforts and aid authorities in managing landslide-prone areas [15,16] by estimating, for example, the correlation between the Precipitation and the occurrence of Shallow Landslides [17].

These methods have been applied in various aspects of landslide prevention, including landslides detection based on images, susceptibility assessment, and the development of warning systems [18]. Recent studies have also focused on the development of dynamic landslide probability maps directly through the application of ML models, indicating the potential for spatiotemporal landslide prediction [19]. ML methods have been recognized for their ability to learn the nonlinear relationship between landslide occurrence and environmental factors, providing a powerful data-driven approach for landslide susceptibility analysis [20].

The use of synthetic aperture radar (SAR) and interferometric SAR has shown promise in detecting and monitoring landslides where intensity and coherence information of SAR images for landslide mapping at a pixel level, demonstrating the potential of SAR for detailed landslide mapping [21]. However, it is important to consider that the success of interferometric SAR for landslide monitoring is contingent on good coherence, which can be challenging in vegetated areas, emphasizing the importance of considering environmental factors in SAR-based landslide forecasting [22]. These studies collectively underscore the potential of SAR and interferometric SAR in landslide detection and monitoring, while also highlighting the challenges associated with coherence in vegetated areas. The use of SAR in landslide forecasting can provide valuable insights into the spatial and temporal dynamics of landslides, contributing to improved hazard assessment and early warning systems [23].

Overall, the literature supports the effectiveness of ML in landslide forecasting and susceptibility mapping, with various studies demonstrating the potential of ML models in improving prediction accuracy, spatial analysis, and dynamic forecasting of landslides. Based on the provided references, several studies offer valuable insights into landslide forecasting and risk assessment, which can be relevant to the context of landslides in Ometepe Island, Nicaragua.

In addressing the pressing issue of landslides in Ometepe, our study confronts several critical challenges. The region's high susceptibility to earthquakes and the sporadic satellite coverage (once every two weeks) compound the difficulty in forecasting. Furthermore, the correlation of extreme rainfall events with hurricanes adds another layer of complexity. Our research aims to tackle these challenges, providing crucial insights into landslide prediction, a matter of significant importance for the region's safety and resilience. It not only navigates through the complexities of natural phenomena but also highlights the vital need for improved forecasting in disaster-prone areas.

Ometepe is a tropical island covered by dense vegetation and subject to high precipitation levels. This research explores the use of Clustering analysis for the coherence Layer to estimate landslide prediction at the level of pixels as an alternative for interferograms with noise from vegetation [22].

The novelty of this work lies in the use of Clustering Analysis of the Coherence Layer coming from InSAR imagery combined with the application of machine learning techniques for landslide forecasting. Unlike traditional methods that heavily rely on physical models, this study leverages data-driven models, employing machine learning to analyze and predict landslides. A key aspect of this research is its dynamic analysis of remote sensing data, focusing on the sensitivity of movements

under various thresholds. This approach allows for a more nuanced understanding of landslide dynamics, especially in the context of hydrological triggers. The use of machine learning to calibrate models based on the availability and characteristics of remote sensing data offers a novel perspective in landslide prediction, addressing the gap left by the limitations of conventional remote sensing methods.

This study evaluates various coherence thresholds against the spatio-temporal aggregation of precipitation from the preceding seven days to examine the compounding effects and influence of continuous rainfall in the area. Three machine learning methods were employed to analyze these variables, aiming to determine the optimal coherence threshold and model for predicting shallow landslide occurrences in this dataset.

This approach comprises three main phases: firstly, the creation of a database incorporating InSAR images and spatio-temporal precipitation, exploring coherence as a potential predictor of shallow landslide occurrence. Secondly, the implementation of machine learning models—specifically Logistic Regression (LR), Random Forest (RF), and Support Vector Classifier (SVC)—to identify the coherence threshold that yields the highest accuracy for predicting shallow landslides using precipitation as an input variable. Finally, the analysis and comparison of the performance of these machine learning models to determine the most effective method for correlating precipitation with shallow landslide occurrence in this specific dataset.

1.1. Case Study

This study focused on Ometepe Island (refer to Figure 1), situated in the center of Lake Cocibolca, the largest freshwater lake in Central America. Spanning an area of 276 km^2, the island was formed through the volcanic activity of Concepción and the now-extinct Maderas Volcanoes. Presently, the entire population is exposed to the direct or indirect influence of Concepción Volcano. However, due to its location within Lake Cocibolca, the island faces constraints in terms of response actions for emergencies, such as the need for quick and effective evacuation.
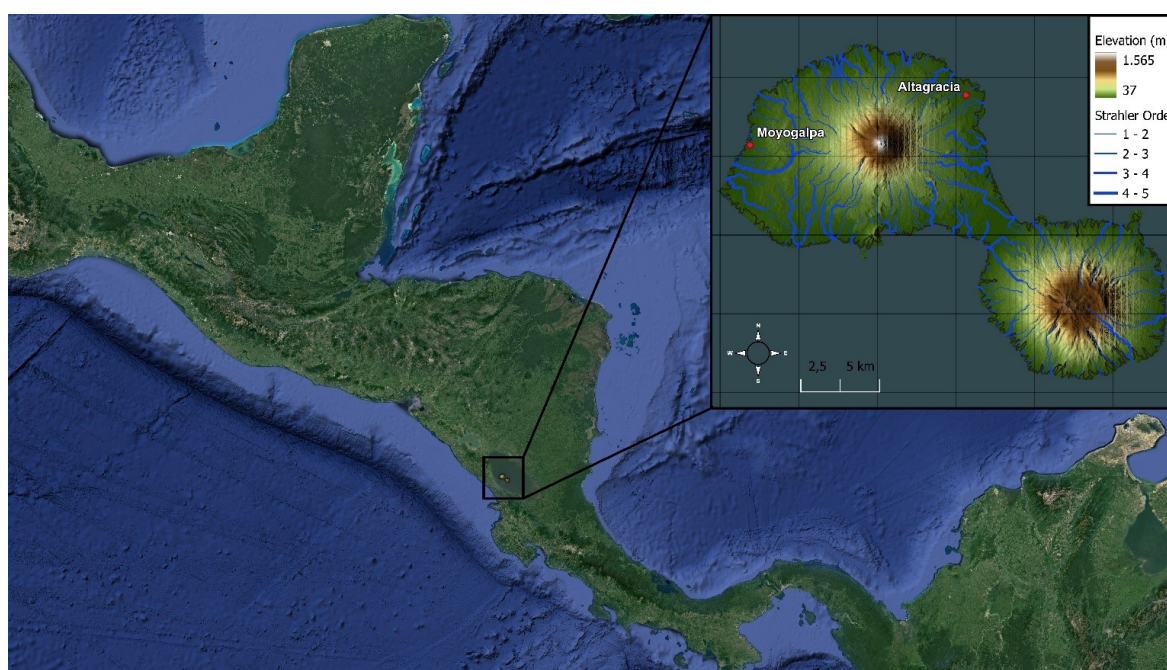

Figure 1. Ometepe Island Location.

Ometepe comprises two municipalities situated in close proximity to Concepción Volcano. The first, Altagracia, is located to the northeast of Concepción Volcano (CV), with a population of 19,995 inhabitants. Within Altagracia, 79.38% reside in 20 villages, while the remaining 20.62% constitute

the urban population. The second municipality is Moyogalpa, situated to the west of CV, with a total population of 9,729. Here, 29.85% of the population is distributed across eight villages, and the urban population represents 70.15% of the total [24]. Presently, the primary socioeconomic activities on the island include agriculture (corn, melon, and bananas), fishing, and livestock [25].

1.2. Topography and Soil Properties

The origin of Ometepe Island can be attributed to the volcanic activity of Concepción and Maderas Volcanoes, resulting in two primary geological formations: Volcanic Pyroclasts and recent sedimentary deposits. Volcanic pyroclastics consist of a mixture of lapilli, ashes, and coarse material. In drier conditions near Concepción Volcano, these formations tend to give rise to black soils with loamy and loamy-sandy textures. On the other hand, recent sedimentary deposits are composed of materials carried by water currents from the upper regions of Concepción Volcano. In the lower areas of Concepción Volcano, black soils with a loamy texture are formed, while in proximity to the lake, the deposits transform into brown-and-clay soils. These soils undergo complete sealing during the wet season and cracking during the dry season due to the presence of smectite [26]. Despite attempts to create a landslide database, available information is limited [27].

1.3. Hydrology

The precipitation data were obtained from The Climate Hazards Group Infrared Precipitation with Stations dataset [28], which has a resolution of 0.05 degrees. Ometepe Island is situated in the dry corridor of Central America, comprising tropical dry forest areas. The precipitation on the island varies from 1000 to 1845 mm/year. It is characterized by rapid and intense rainfall, followed by periods with reduced or no precipitation, as illustrated in Figure 2:
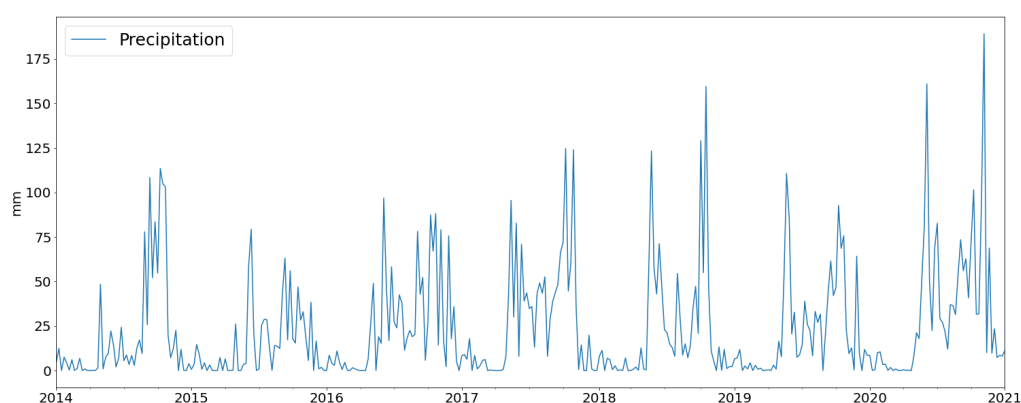

Figure 2. Ometepe Island Precipitation.

Ometepe experiences a tropical climate characterized by two distinct seasons: wet and dry. The wet season extends from May to November, while the dry season spans from December to April. This seasonal pattern is evident in Figure 3, where the wet season displays a notable decrease in precipitation during July and August. Consequently, the majority of precipitation is concentrated in two distinct periods: May–June and September–October.

Due to significant noise (incoherence) in the images of the Maderas Volcano area during the SAR analysis and considering that the majority of the population resides around Concepción Volcano, whose catchment differs entirely from Maderas Volcano, this study is concentrated on the Concepción Volcano area. The yearly average of spatially distributed precipitation for the study area is depicted in Figure 4.
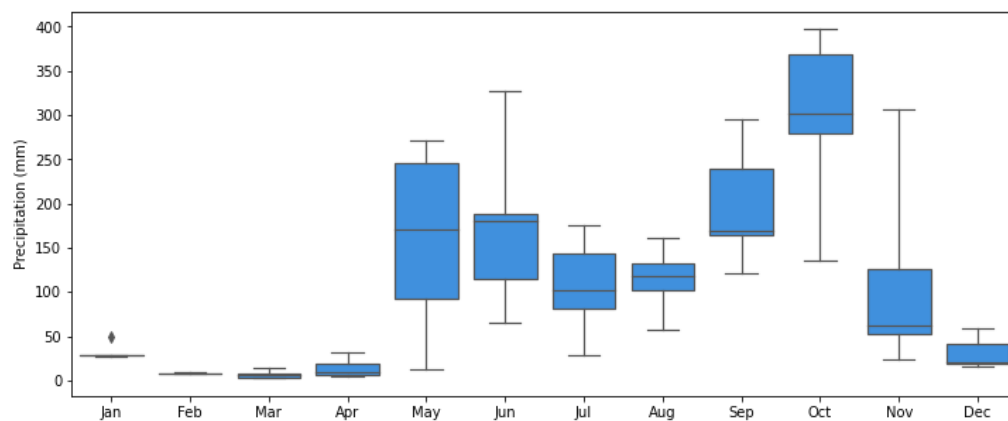
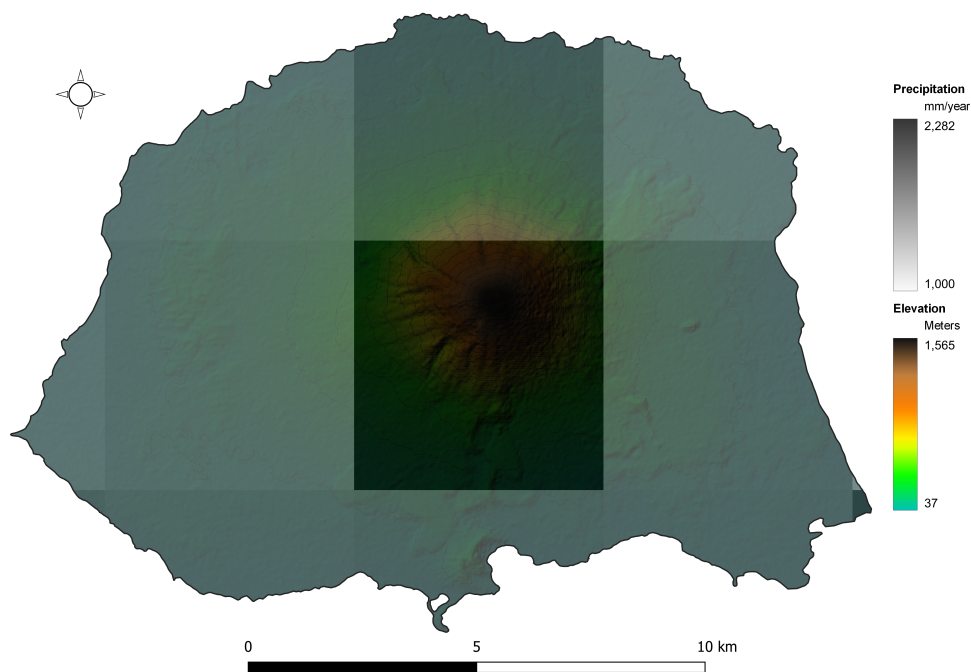**Figure 3.** Average Monthly Precipitation of Ometepe Island.

Figure 4. Spatial Precipitation Ometepe Island.

As evident in the figure, the highest precipitation levels, reaching 2282 mm/year, are observed in the upper parts of the island, near the summit of the volcano.

1.4. SAR Analysis

Remote sensing plays a crucial role in landslide forecasting and monitoring, offering valuable insights into environmental factors that contribute to landslides. Various remote sensing techniques, such as Synthetic Aperture Radar (SAR), have been utilized for landslide detection and classification [29–33]. These technologies provide a comprehensive view of the terrain and can capture changes in land surface conditions that may indicate landslide susceptibility.

Sentinel-1 satellite data, with its Synthetic Aperture Radar (SAR) potential, has been instrumental in landslide detection, monitoring, and assessment due to its all-weather imaging capabilities and high spatial resolution [34]. Sentinel-1 Single Look Complex (SLC) data contains information on both the amplitude and phase of the radar signal, enabling detailed analysis of radar backscatter characteristics [35]. The SLC data is used in conjunction with other satellite data sources to monitor phenomena

like glacier movements and land deformations, showcasing its versatility in different environmental monitoring applications [36,37].

An interferogram is a crucial product of Synthetic Aperture Radar (SAR) interferometry, providing valuable insights into surface deformations and changes over time. It is generated by measuring the phase difference between two SAR images acquired at different times, allowing for the detection of ground movements with high precision [38–41]. These interferograms are essential for monitoring phenomena such as landslides, volcanic activities, and glacier movements [42–44]. The interferometric correlation maintained in SAR imagery enables the measurement of ground displacement with high precision, making it a valuable tool for geodetic studies [43,44].

The coherence value indicates the similarity between the radar signals received during the two image acquisitions. The coherence of interferograms is essential for identifying coherent pixels and extracting meaningful deformation information. It is used to assess the reliability of the interferometric results and to filter out noisy or decorrelated areas, ensuring that the deformation measurements are accurate and trustworthy [45]. By analyzing the coherence of interferograms, researchers can evaluate the quality of the data, identify areas of interest for further study, and enhance the interpretation of surface deformation patterns [46].

However, it is important to note that dense vegetation may affect the penetration of the SAR waves into the ground [47]. Nevertheless, recent studies have focused on utilizing other SAR-derived information like amplitude, phase and coherence for landslide mapping and detection, showcasing the importance of coherence analysis in landslide studies with complex SAR interactions [48,49].

Lately, studies have explored and shown the effectiveness of remote sensing in landslide detection using machine learning algorithms applied to satellite imagery [32,33]. These studies successfully applied machine learning models like Random Forest, Boosted Regression Tree, Logistic Regression, and Artificial Neural Networks for landslide susceptibility mapping, demonstrating the effectiveness of these algorithms in analyzing landslide-prone areas [13,50].

In conclusion, remote sensing technologies offer a powerful tool for landslide forecasting, detection, and monitoring. This study explores the use of Sentinel-1 satellite SAR imagery and spatiotemporal precipitation data for landslide forecasting using three machine learning approaches to classify shallow landslide events. The Methodology used for this reasearch is explained next.

2. Methodology

Figure 5 shows an overview of the process followed in this research. Three main stages were established: the first was Database Creation, where the input and output variables were generated from Remote Sensing data; the second was Machine Learning, where three models were created using LR, RF and SVC to identify the best coherence threshold for Shallow Landslide Occurrence (SLO); and the third was Analysis of Results, where all models were compared to each other to determine the best ML method for this case. These three Phases are described below:
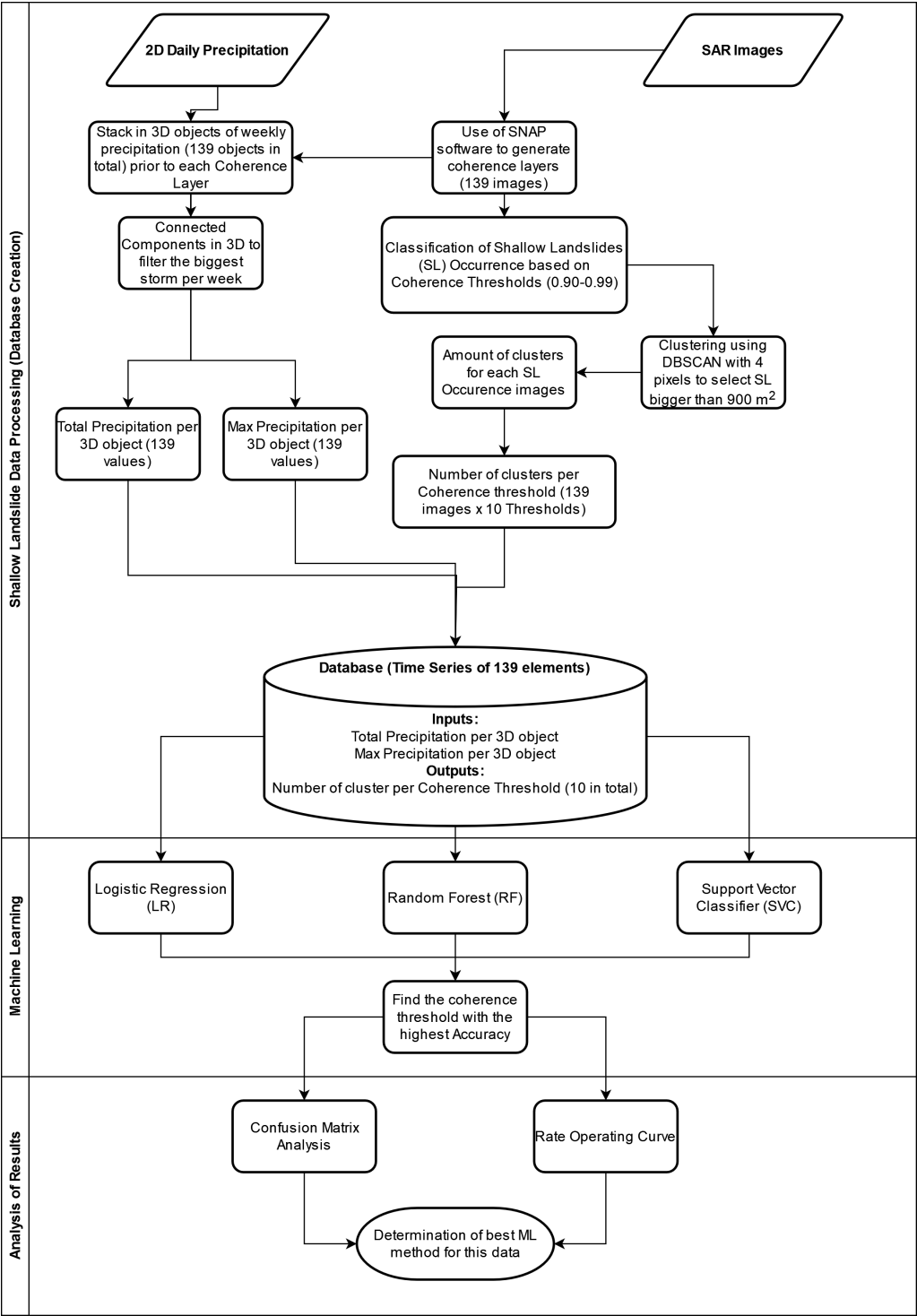
Figure 5. Research Steps for Shallow Landslide Analysis.

2.1. Phase 1: Shallow Landslide Data Processing

2.1.1. Overview of SAR Images Processing

In Figure 6, the steps undertaken to create the Shallow Landslide Occurrence Maps (SLOM) from the SAR images spanning 2015–2020 are illustrated. The initial step involved loading two SAR images from the Sentinel-1 Satellite into the SNAP software, developed by the European Space Agency. The first process entailed co-registering both images, with the second sub-swath (IW2) chosen for its coverage of the entire island. The Orbit State Vector used was Sentinel Precise, with a Polynomial

Degree of 3. For the geocoding subprocess, the Digital Elevation Model (DEM) from the Alaska Satellite Facility (ASF) with a resolution of 12.5m was utilized, employing bilinear interpolation for both DEM resampling method and type.
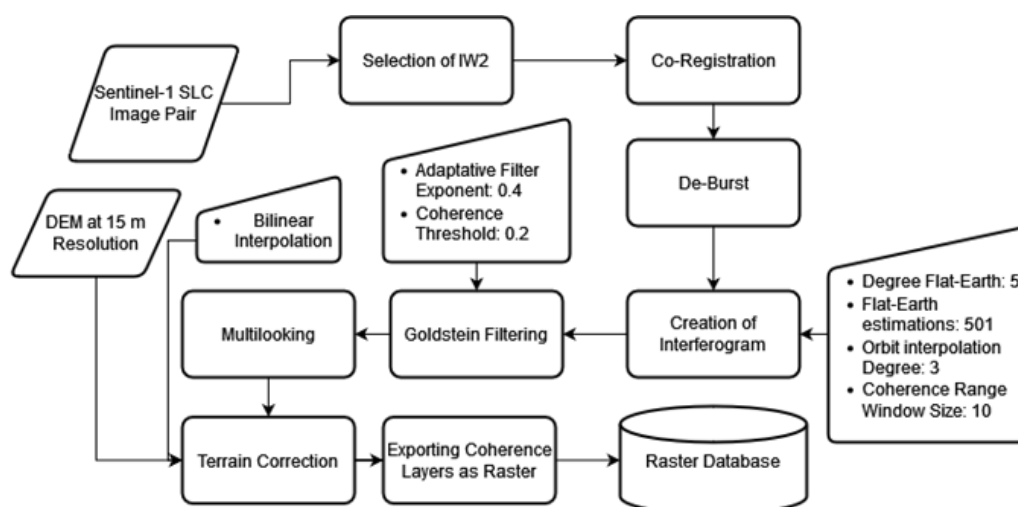

Figure 6. SAR Images Steps.

Following the successful co-registration of the image pair, the de-bursting process was initiated. This process involved merging the stripes of the IW2 for unification into a single image. Subsequently, the next step involved the creation of the interferogram with the following parameters:

- Degree of flat-earth phase: 5
- Number of flat-earth estimation points: 501
- Orbit interpolation degree: 3
- Coherence range window size: 10

After the creation of the interferogram, a Goldstein Filter was applied with an adaptive filter exponent of 0.4 and a coherence threshold of 0.2. This was followed by the multi-looking phase, where pixels were compressed to uniform dimensions (four times the number of columns in relation to the rows). The final step involved applying terrain correction to the interferogram, using the DEM from ASF for adjustment and setting it to a resolution of 15 by 15 m using bilinear interpolation. The resulting file, the Coherence Layer, was then exported for further analysis.

2.1.2. SAR Images for Shallow Landslide Occurrence

The images utilized in this study were Sentinel-1 satellite (SAR) images in Single Looking Complex (SLC) and IW format, spanning from the beginning of 2015 to the end of 2020, with a 15-day interval between acquisitions. In total, 146 images of the island were included in the analysis. Each image (Im) has a resolution of n x m, as illustrated in Equation 1:

$$Im_t = Pixels_{nxm_t} \quad where \quad t = \{1, 2, 3...146\} \tag{1}$$

The concept of coherence in an image serves as an indicator for shallow landslides (Events). Coherence, as defined in Equation 2, quantifies the similarity between two images in terms of both phase and intensity. When the similarity is high, the coherence value approaches 1; conversely, a low similarity results in a coherence value close to 0. It is essential to note a delicate balance: there exists a boundary between differences in coherence that signify topographic changes on the territory and noise. Occasionally, coherence values (similitude) can be influenced by environmental interference. In this study, Coherence Thresholds (CT) ranging from 0.90 to 0.99 were tested to classify the occurrence of an actual shallow landslide. Pixel values exceeding the threshold were set to 1 (Event), while those below were set to 0 (No Event), generating the Shallow-Landslide Occurrence Maps (SLOMs).

$$Coh_i^t = f(Im_i, Im_{i-1}) \quad where \quad t = \{1, 2, 3...145\} \tag{2}$$

After generating the 145 Shallow-Landslide Occurrence Maps (SLOMs) using SNAP software and Python, the pixel resolution was set to 15 m. These SLOMs were then imported into a NumPy Array with the objective of filtering out images that were not at least seven days apart. This seven-day threshold represents the temporal influence of precipitation on Shallow Landslide Events. Following the removal of images with less than seven days between them, the outcome was 139 SLOMs covering the period from 2015 to the end of 2020, as indicated in Equation 3.

$$SLOM = f(Coh)_{nxm}^t \quad where \quad t = \{1, 2, 3...139\} \tag{3}$$

At this stage, there were 139 Shallow-Landslide Occurrence Maps (SLOMs) represented as cells with values of 0 or 1 stacked on the Z-axis in NumPy. The subsequent step involved implementing the Density-Based Scan Algorithm to identify clusters of event occurrences (Shallow Landslides) in the SLOMs. This process aimed to eliminate noise arising from isolated pixels with values of 1 and identify adjacent cells of a minimum size. The Density-Based Scan Algorithm was employed for this purpose.

Density-Based Scan Algorithm

The Density-Based Scan Algorithm is employed to identify clusters in data characterized by high-density spots. Clusters detected by DBSCAN can take any shape, as the algorithm relies on the concept of Core Samples, denoting clusters of points in close proximity to each other in terms of space. A cluster is defined as a collection of points within a specified radius. The algorithm takes two parameters: the radius (Epsilon) and the minimum quantity of points (minimum samples). During evaluation, each point in the dataset counts the number of points within the specified radius. All points are considered Core Samples by definition. Any point not classified as a Core Sample and lacking neighbors within the specified radius is deemed an outlier [51].

DBSCAN typically operates with vector data rather than matrices (cells). To accommodate this, the pixels representing events in each Shallow-Landslide Occurrence Map (SLOM) were converted into coordinates, utilizing their relative positions within the matrix (n x m). Subsequently, the indices of the position of each cell (row and column) were used as coordinates, serving as input for the DBSCAN process. This step facilitated the removal of points (coordinates) that did not meet the clustering criteria, including the minimum samples (minimum cluster size) and radius (distance between points).

In this case, the radius was kept constant because the distance between cells on the Shallow-Landslide Occurrence Map is always 1 or $\sqrt{2}$ when the cell is diagonal (refer to Figure 7). Consequently, a constant value of 1.5 was applied. This led to the removal of coordinates that did not meet the clustering criteria. Subsequently, these coordinates were reintroduced into the Shallow-Landslide Occurrence Map (SLOM) to identify which pixels belonged to a cluster, while the rest were set to 0 as they represented noise in the sample. This process was repeated for all SLOMs (a total of 139) and stacked together to create the SLOM database. The SLOM, represented as a 2-dimensional array of ones and zeros, was inputted into the algorithm, which counted and labeled the clusters based on the specified radius and minimum samples.

Finally, a time series consisting of 0s and 1s was derived, indicating the occurrence of at least one event in the SLOM. This time series, composed of 139 elements, served as the output for use in the machine learning analysis.
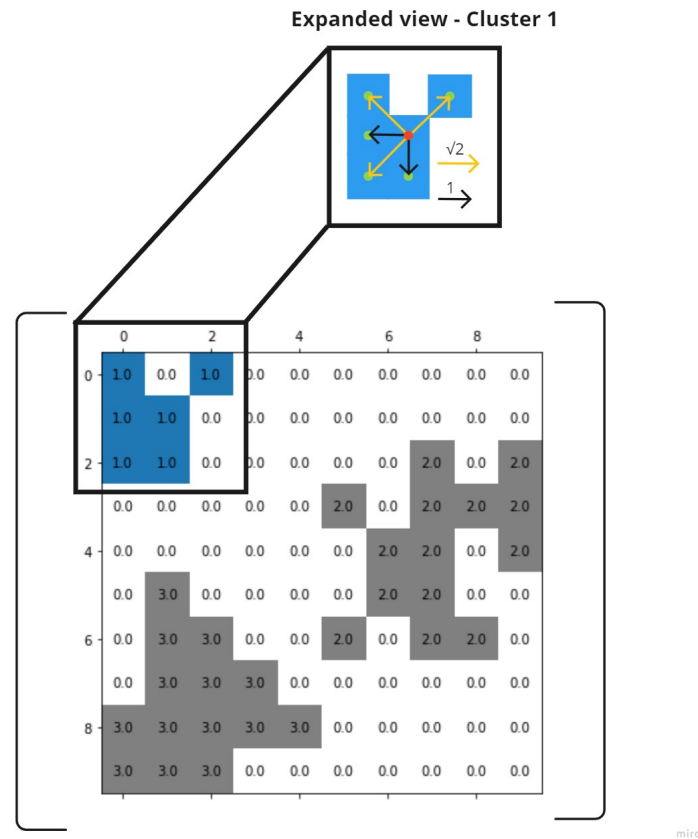
Figure 7. DBSCAN - Visual Representation of SLOMs.

2.1.3. Spatiotemporal Precipitation

Hydrometeorological variables influencing landslides encompass precipitation, soil moisture, capillarity forces, recharge pressures, surface runoff, and others. However, this work specifically focuses on considering precipitation and its spatial and temporal variations as indicators of hydrometeorological forcing on shallow landslides. The impact of these variables can be implicitly included by analyzing the temporal distribution and influence of rainfall across the territory.

Precipitation data were sourced from the CHIRPS website, a widely used resource in experimental studies in the region, chosen for its documented results. Daily data spanning from 2015 to the end of 2020 was downloaded, with a spatial resolution of 0.05 degrees (approximately 5 km). This dataset comprised 2557 images of Ometepe Island, where individual cells represented the total daily rainfall (in mm) on the island for a specific day. Each day's data was stored in a separate file (raster). The steps followed in this section were:

Downscale precipitation data into the same scale as the SLOM

The data was loaded into a local file system, and the resolution was adjusted to 15 m through linear interpolation, aligning it with the resolution of the Shallow-Landslide Occurrence Map (SLOM) following the procedure outlined in [52]. Subsequently, the resulting rasters were stored in a 3D NumPy Array [53]. The X and Y dimensions represented the island coordinates, while the Z dimension denoted time (precipitation for that day). This arrangement facilitated the creation of an aggregated Precipitation object, as depicted in Equation 4:

$$O = P_{nxm}^t t = \Sigma i \quad where \quad i = 7 \tag{4}$$

Where O is the 3D precipitation object of dimensions n rows by m columns by t time, where time is equal to 7, as the precipitation objects were considered the influence of seven days of rainfall prior to a SLO.

Data cleansing and Connected-Component Labeling

The Precipitation object was trimmed using the same Shapefile delineating the island's contour, effectively removing precipitation values outside the Study Area. The eliminated values were then replaced with a Mask Value of -99. In total, there were 139 Shallow-Landslide Occurrence Maps (SLOMs). Each SLOM was paired with the respective Precipitation object, resulting in a total of 139 samples. These samples shared the same spatial resolution and represented aggregated precipitation during the 7-day period preceding the date of the corresponding SLOM.

Subsequently, a 3D-connected component analysis was applied to each Precipitation Object, isolating only the largest precipitation cluster based on the total volume of rainfall. In this research, the assumption is that a Shallow Landslide is triggered by the most substantial storm (cluster with the highest precipitation) within the 7-day window preceding the date of the Shallow-Landslide Occurrence Map (SLOM). Consequently, only the largest storm was utilized in creating the model. Each Precipitation Object, paired with its corresponding SLOM, underwent the connected component algorithm [54]. The algorithm labels interconnected pixels with precipitation values greater than 0, determining the number of clusters per 3D matrix. This process was executed along the X, Y, and Z axes, resulting in tri-dimensional precipitation clusters. This procedure was repeated for each SLOM and Precipitation Object, totaling 139 samples.

2.2. Phase 2: Machine Learning

From the spatiotemporal precipitation, two variables were generated: one representing the total volume of precipitation and the other indicating the maximum precipitation value. These variables were used as input for the Machine Learning (ML) methods: Logistic Regression (LR), Random Forest (RF), and Support Vector Classifier (SVC), as the Shallow-Landslide Occurrence Map (SLOM) is influenced by the total volume of the most significant rainfall during the 7-day period.

Out of the 139 samples, two time series were derived: one featuring the normalized total volume of rainfall from the precipitation object, and the second showcasing the normalized maximum precipitation values. These time series served as input variables for the ML process.

The ML methods employed in this study included Logistic Regression (LR), Random Forest (RF), and Support Vector Classifier (SVC). The dataset was split into training and testing sets, with a proportion of 75/25, respectively. To optimize the models' performance, the statistical characteristics of the Shallow-Landslide Occurrence Maps (SLOM) were evenly distributed between the training and testing sets. This ensured that the proportion of events and no events was consistent for both sets (refer to Table 1):

Table 1. Machine Learning Parameters for Each Method.

Logistic Regression (LR)	Random Forest (RF)	Support Vector Classifier (SVC)
C: 0.5-5	N estimators: 1-10	C: 0.1-10
Solver: 'liblinear', 'lbfgs'	Criterion: 'gini'	Kernel: 'sigmoid','polynomial', 'rbf'
Random State : 0-30 000	Random State : 0-30 000	Random State : 0-30 000

2.2.1. Identification of Best Coherence Threshold

Each coherence threshold resulted in different event counts, as expected, since the coherence threshold signifies the sensitivity level of the Shallow-Landslide Occurrence Map (SLOM). A lower coherence threshold corresponds to a higher number of events, and vice versa. Nonetheless, there exists a point where there is maximum correlation between precipitation and the occurrence of these

events, crucial for creating models using machine learning (ML). This can be determined by the following Equation (5):

$$A_R = A_M - A_{NM} \tag{5}$$

Where the real accuracy (AR) is defined as the difference between the Accuracy of the model (AM) and the ratio of events (or no events) with respect to the total of SLOM (139), which ends up being the accuracy of no model. At the same time, the accuracy of the model (AM) depends on Precipitation and the number of Events (E), while the accuracy of the no-model (ANM) is just the ratio of Events (E) out of all SLOM. So, the equation of AR can be redefined as Equation (6):

$$A_R = f(P, E) - \frac{E}{SLOM} \tag{6}$$

The events, on their part, depend on the Coherence (C) and cluster size. As this research focused on events larger than 900 m^2, the cluster size was constant and therefore can be eliminated. Then, the Events become:

$$E = g(C) \tag{7}$$

The events, on their part, depend on the Coherence (C) and cluster size. As this research focused on events larger than 900 m2, the cluster size was constant and therefore could be eliminated. Then, the Events become:

General Equation for Event Ratios >= 0.5

$$A_R = f(P, g(C)) - \frac{g(C)}{SLOM} \tag{8}$$

General Equation for Event Ratios <= 0.5

$$A_R = f(P, g(C)) + \frac{g(C)}{SLOM} - 1 \tag{9}$$

With Equation (8) and Equation (9), it was possible to create the ML models and determine which Coherence Threshold was the best-performing.

2.3. Phase 3: Analysis of Results

After finding the best coherence threshold, the next step was to take a deeper look into each of the ML models to determine the best method for this specific data. As this is a binary-classification problem, the following classifiers were used:

- True Positive (TP): when the outcome of the model (prediction) is positive and the actual result is also positive
- False Positive (FP): when the outcome of the model (prediction) is positive and the actual result is also negative
- False Negative (FN): when the outcome of the model (prediction) is negative and the actual result is also positive
- True Negative (TN): when the outcome of the model (prediction) is negative and the actual result is also negative

From here, Confusion Matrices were generated for each ML model to have a more insightful perspective; this was accompanied by ROC curve graphs to better understand the stability and performance of the LR, RF and SVC models.

Finally, these classifiers were used to generate the following indices to measure the performance of the model [55]:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

$$FPR = \frac{FP}{TN + FP} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F_1 score = \frac{2(Precision)(Recall)}{Precision + Recall} \tag{15}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{16}$$

$$Overall\,Accuracy = Accuracy + F_1 + MCC \tag{17}$$

Where Accuracy (Equation (10)) measures the ratio of correct measurements for the binary classification; Specificity (Equation (11)) measures the ratio of TN amongst all negative samples (including FP); False Positive Rate (FPR) from Equation (12)indicates the ratio of negative samples that were mistakenly classified as FP; Precision (Equation (13)) is the ratio of TP amongst all positive-classified samples (including the ones mistakenly classified); Recall (Equation (14)) is the ratio of TP amongst all positive samples, including the ones incorrectly classified as FN; F1score (Equation (15)) is the harmonic mean between precision and recall; The Mathews Correlation Coefficient (MCC) from Equation (16) is a measure of the quality classification of two classes (Matthews, 1975); The overall Accuracy of the model (Equation (17)) is just the sum of the MCC, Accuracy and F1 score. This Overall Accuracy index was the determinant variable for evaluating the performance of the LR, RF and SVC models. All these results, from the Creation of the Database (Phase 1) to Machine Learning (Phase 2) and Analysis of Results (Phase 3), are displayed next.

3. Results and Discussion

3.1. Phase 1: Shallow Landslide Data Processing

3.1.1. SAR Images

The initial step in the Shallow Landslide Processing involved generating Coherence Maps as rasters with a resolution of 15 m. Figure 8 displays the Intensity and Coherence layers resulting from the Interferogram creation. There is significant interference across the island, making it nearly indistinguishable from the water of Lake Cocibolca. Hence, the coherence layer served as an indicator of Shallow Landslide Occurrence (SLO), following the steps outlined in Phase 1: Shallow Landslide Data Processing. Subsequently, the variation in SLO will be presented when different Coherence Thresholds (CTs) are applied to each Shallow Landslide Occurrence Map (SLOM).
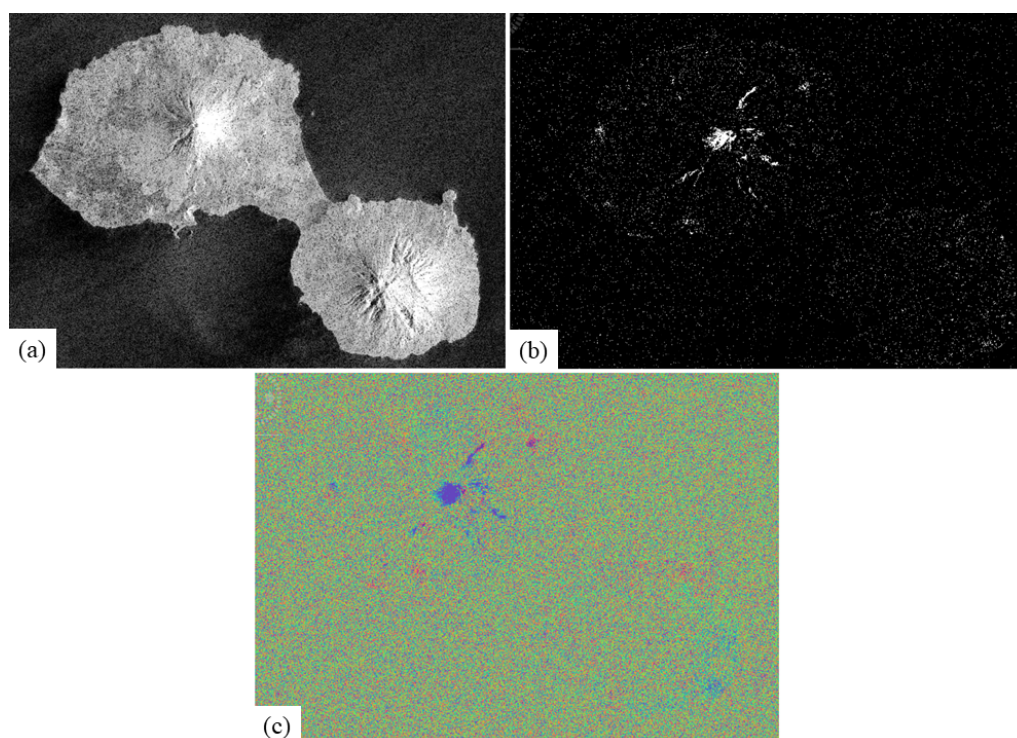
Figure 8. Visualization of the SAR Layers (**a**) Intensity. (**b**) Coherence. (**c**) Interferogram.

3.1.2. SAR Images for Shallow Landslide Occurrence

For the generation of the SLOM, CT ranging from 0.9 to 0.99 were tested. The DBSCAN Clustering Algorithm was used to identify the clusters (Shallow Landslides) with an area of at least 4 pixels, which is roughly 900 m^2. Below is showed how the differences in CT lead to different numbers of events (clusters):

Figure 9 illustrates the number of clusters per sample, where a lower threshold indicates reduced sensitivity to noise, leading to a higher number of clusters per image. It's crucial to note that as the coherence threshold decreases, the number of clusters increases, reaching a point where events are consistently detected. Conversely, higher thresholds, exceeding 0.97, result in almost no occurrence of events. Beyond this threshold, events are not effectively represented. This observation prompts further examination of the relationship between the coherence threshold and the number of clusters.
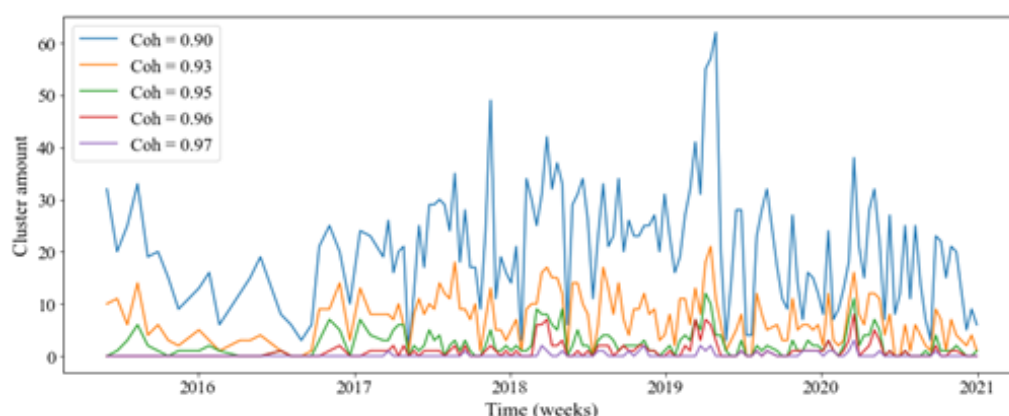

Figure 9. Cluster Amount for each SLOM at different CT.

3.1.3. Spatiotemporal Precipitation

The total precipitation volume for the most significant storm within the 7-day window corresponding to the date of each Shallow Landslide Occurrence Map (SLOM) is presented below (see

Figure 10). There is a noticeable difference in precipitation between the dry and wet seasons (May to October and November to April, respectively), with a peak observed in 2018.
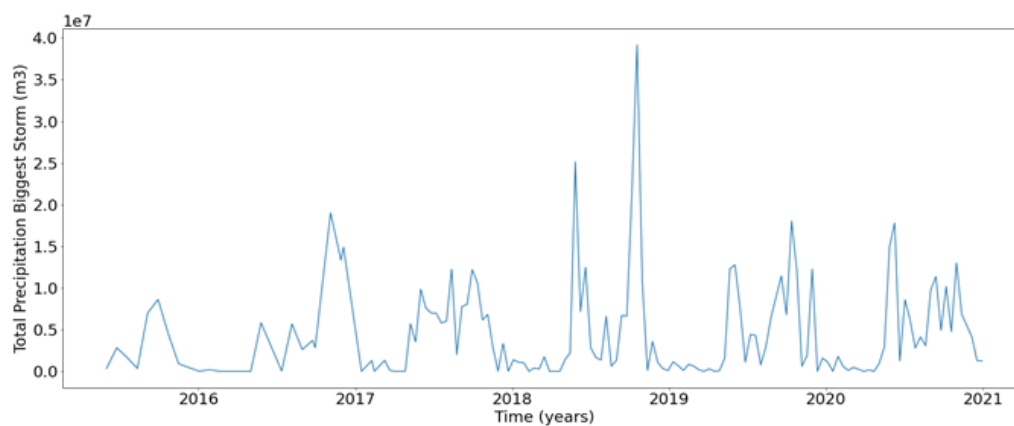

Figure 10. Total Precipitation of Biggest Storm.

Besides the total precipitation time series, another input variable was derived from the spatially distributed precipitation: the maximum daily precipitation. Following the same methodology as the 7-day window, the maximum precipitation value was extracted from the cells, generating another time series (refer to Figure 11):
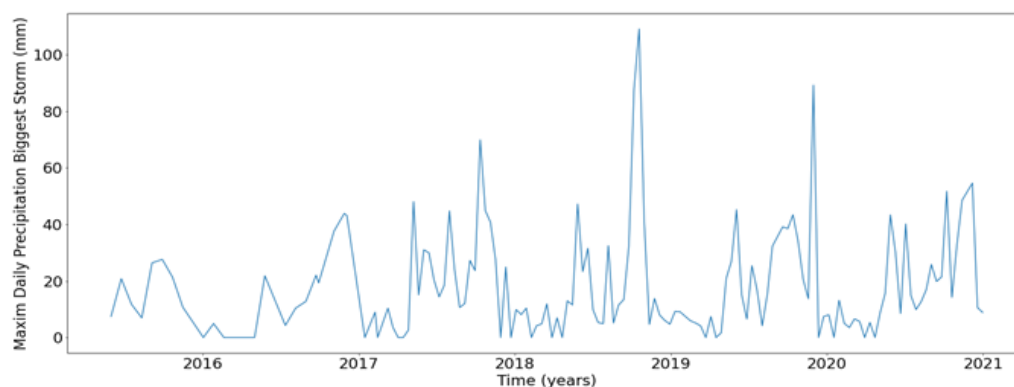

Figure 11. Total Precipitation of Biggest Storm.

It is important to note that the behavior is quite similar to the total volume. For this reason, both input variables were plotted against each other, as displayed in Figure 12:

Most points on the scatter plot are above the diagonal (slope of 1), indicating that the storms were of short duration and high intensity, a factor that can trigger Shallow Landslide Events.
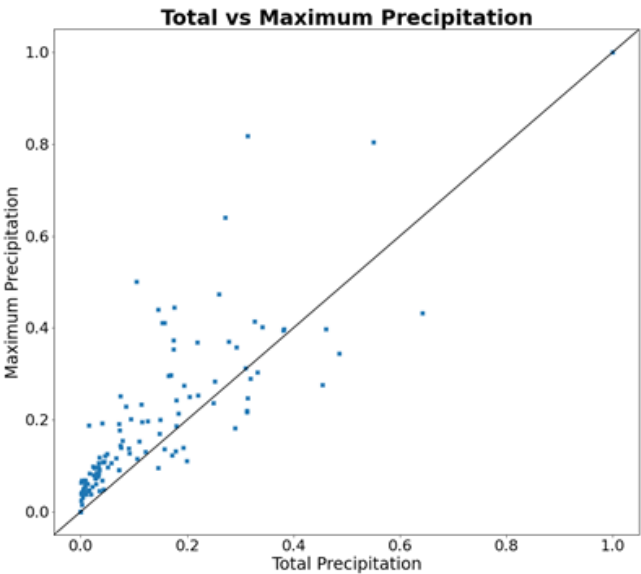
Figure 12. Total Precipitation of Biggest Storm.

3.2. Phase 2: Machine Learning

3.2.1. Model Creation

After testing different criteria for creating the models using the three methods described in Methodology, LR, RF & SVC, the final models were created using the following parameters showed on Table 2:

Table 2. Machine Learning Parameters for Each Method.

Logistic Regression (LR)	Random Forest (RF)	Support Vector Classifier (SVC)
C: 1	N estimators: 5	C: 1
Solver: 'liblinear'	Criterion: 'gini'	Kernel: 'sigmoid'
Random State: 2274	Random State: 2516	Random State: 4931

3.2.2. Model Accuracies per Coherence Threshold

Figure 13 describes the variation in CT leads to different numbers of clusters (Events). This is key for determining the correlation between precipitation and Shallow Landslide Occurrence (SLO) using ML. Below is showed the variation in the number of events per CT:
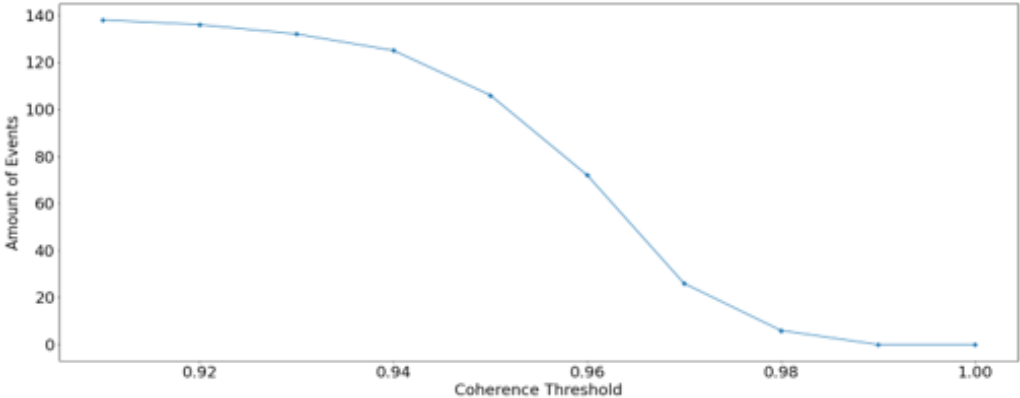

Figure 13. Total Precipitation of Biggest Storm.

As showed in Figure 9, The variation of 0.01 in CT has a significant impact on the number of events, where low coherence thresholds like 0.9 to 0.94 result in events in all (or almost all) SLOM. On the contrary, with high coherence thresholds like 0.98 to 1.00, the SLOM results in no (or almost no) events. The SLOM is sensitive to the CT and could tend to generate more (or fewer) events. The thresholds 0.95, 0.96 and 0.97 produce the most even distribution in terms of the proportion of events and no events. The performance of the model depends not only on the precipitation data but also on the coherence threshold, and the bigger the difference between an event and no event, the worse the model would perform.

This led to first determining which CT had the most correlation to the precipitation data. For this, the three ML methods were used (LR, RF and SVC), with the methodology explained previously to determine the real accuracy of the models, considering the "No-Model", which is the ratio of events and no events in the SLOM. The results are showed below:

As it can be noted in Figure 14, the highest performance of the models is on CT = 0.96, where the proportion of Events and No Events is close to 1 (Event ratio close to 0.5), where SVC had the highest accuracy, followed by RF and LR. It can be noted that the accuracy of the scenarios where the Event ratio is not close to 0.5 is almost equal to the event ratios themselves, which implies that the models are not working or giving useful information and are close to or equal to 0. However, when the coherence threshold gives a ratio close to 0.5, the algorithms perform better. This may be why the behavior was quite similar for all three ML methods; for this reason, a deeper look into each ML method was taken, with CT = 0.96 (as is explained in the following section).
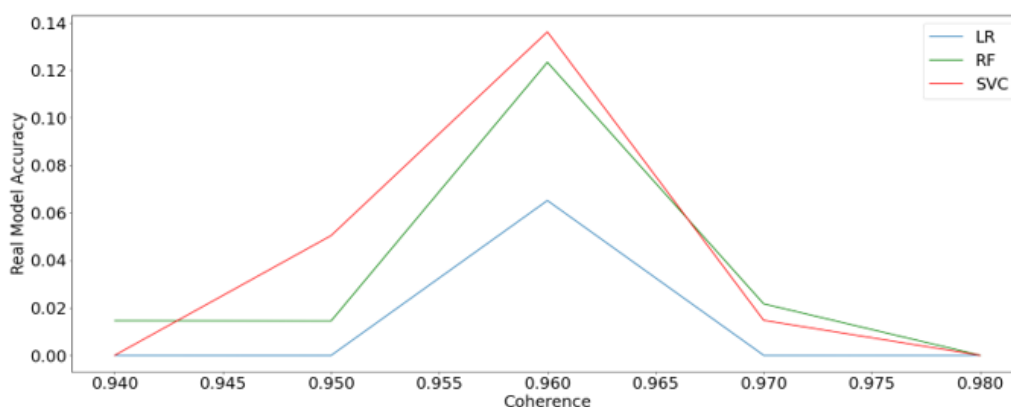

Figure 14. Total Precipitation of Biggest Storm.

3.3. Phase 3: Analysis of Results

3.3.1. Logistic Regression (LR)

The first method that was tested was Logistic Regression. For a coherence of 0.96, it can be seen in Figure 15 that the accuracy was 0.58 and that the division between event and no event made with the LR cannot accurately divide the occurrence of events, tending to generate more FP than FN. This is reinforced by the low percentage of both false positives and true negatives, where both are quite low compared to TP and FP.

Now, in terms of the ROC Curve, LR had a low AUC (0.6). Moreover, the standard deviation from the model (0.05) suggests that there was not much variability between the runs. Now, these results were compared to RF and SVC:
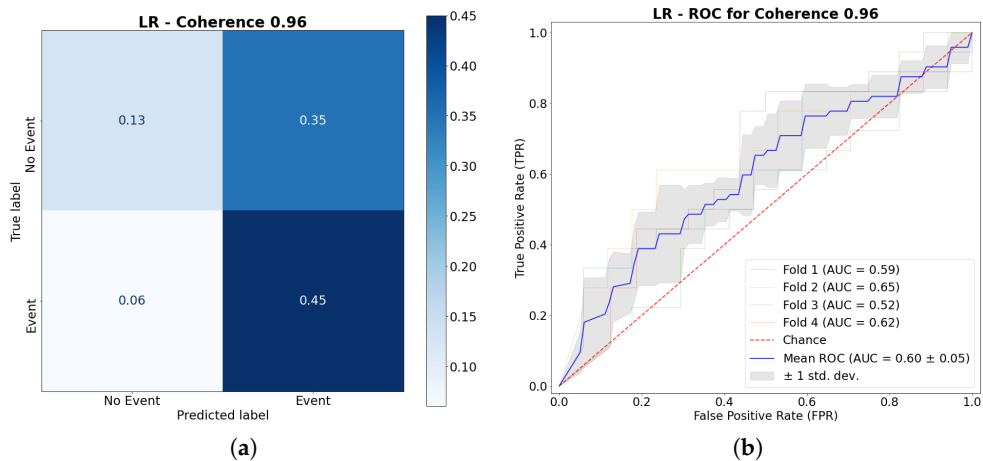
Figure 15. Logistic Regression Analysis. (**a**) Confusion Matrix. (**b**) ROC Curve.

3.3.2. Random Forest (RF)

Figure 16 showcases the confusion matrix of the RF model with a coherence threshold of 0.96. It can be noted that RF correctly identified the occurrence (or no occurrence) of the events (accuracy of 0.64), while also generating similar ratios of FP and FN, meaning the model does not tend to generate one over the other.

Figure 16. Random Forest Analysis. (**a**) Confusion Matrix. (**b**) ROC Curve.

It is important to remark that the standard deviation, that is showcased on Figure 16, is bigger compared to LR's, which means that each decision tree, for each run, produces different results from each other, resulting in higher variability in results. For this reason, the AUC was lower (0.58) than the results from LR (0.60).

3.3.3. Support Vector Classifier (SVC)

Figure 17 showcases the confusion matrix of the SVC model with a coherence threshold of 0.96. It can be noted that SVC correctly identified the occurrence (or no occurrence) of the events (accuracy of 0.66), while also generating equal ratios of FP and FN, meaning the model does not tend to generate one over the other.
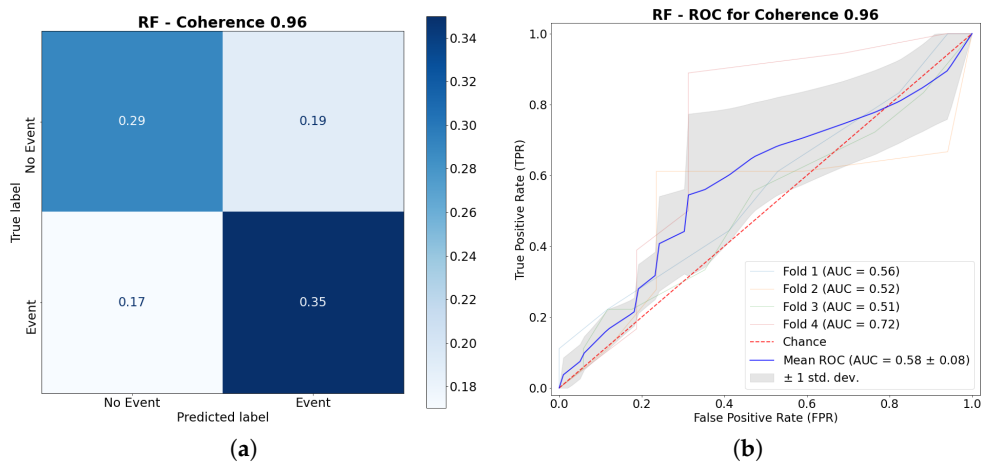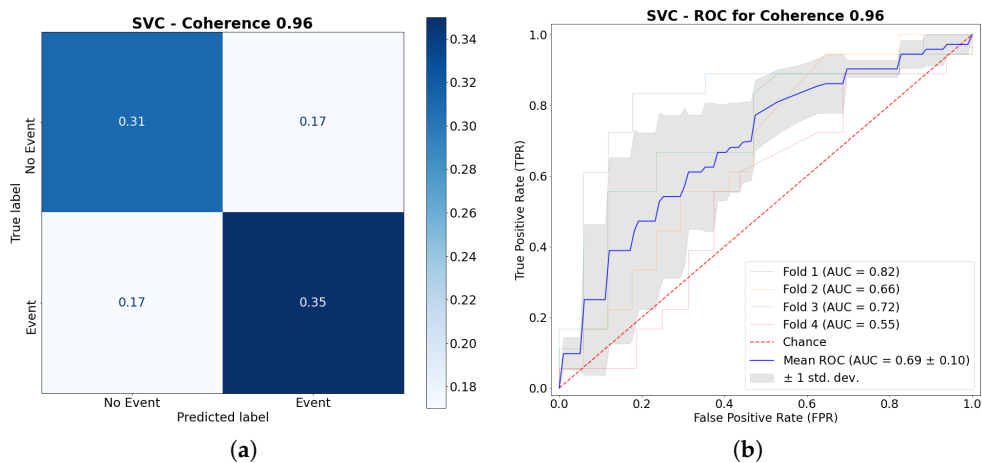
Figure 17. Support Vector Classifier Analysis. (**a**) Confusion Matrix. (**b**) ROC Curve.

It is important to remark that the standard deviation that is showcased in Figure 17 is the biggest of the three models (0.10), which means that for each run, it produces different results from each other. Nevertheless, SVC managed to get a better score of AUC (0.69), which means that despite the high variability, the scores of the model do tend to perform better than LR and RF. From the analysis of accuracy, confusion matrices and ROC curves, SVC was the better performing model because it had the highest accuracy (0.66) and AUC (0.69) despite having the highest standard deviation (0.10).

3.3.4. Best Machine Learning Model

After running the three ML methods and determining the best CT, it was possible to further analyze when the coherence was 0.96 using the indices defined on Methodology. The results are described by the Table 3 with a summary for all ML models with the best CT:

Table 3. Machine Learning Performance Indices.

Indices	Logistic Regression	Random Forest	Support Vector Classifier
TP	0.45	0.35	0.35
FP	0.35	0.19	0.17
FN	0.06	0.17	0.17
TN	0.13	0.29	0.31
Accuracy	0.59	0.64	0.66
Specificity	0.27	0.60	0.65
FPR	0.73	0.40	0.35
Precision	0.56	0.65	0.67
Recall	0.88	0.67	0.67
F1score	0.69	0.66	0.67
MCC	0.19	0.28	0.32
OA	1.47	1.58	1.65

In addition to Table 3, the results are also displayed on Figure 18 to provide a visual interpretation on how each ML method performed against each other:

Starting with Accuracy, all three models performed similarly, with SVC getting the highest value (0.66), indicating that all three models (LR, RF and SVC) are comparable (0.59, 0.64 and 0.66) for correctly identifying "Events" and "No Events".
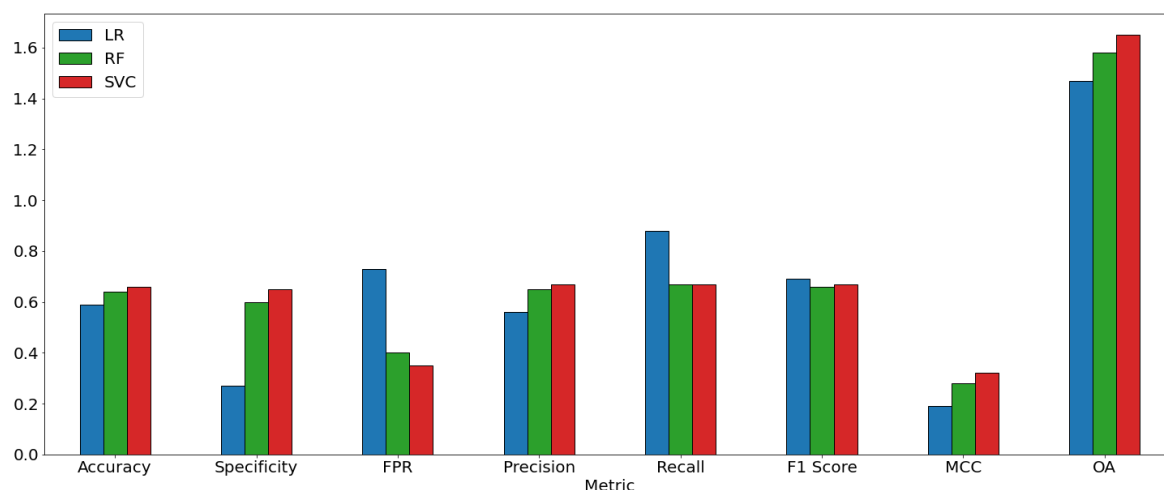
Figure 18. Machine Learning Models Comparison.

Specificity, which measures the ratio of TN amongst all negative samples (TN + FP), was the highest with SVC (0.65), followed closely by RF (0.60), which means that both RF and SVC correctly identified a bigger proportion of "No Events" (Negative Samples) compared to LR (0.27).

In terms of FPR (False-Positive Rate), the ratio of Negative Samples that were incorrectly classified as FP was the highest at LR (0.73), indicating that it tends more to incorrectly classify "No Events" (Negative Samples) compared to RF (0.40) and SVC (0.35).

Precision, the ratio of TP amongst all positive-classified samples, was obtained by SVC (0.67), followed again by RF (0.65), and LR at last (0.56), indicating that SVC and RF have a slight advantage over LR in providing a higher ratio of confidence that a Positive Sample is an "Event".

Recall, that indicates the ratio of "Events" that were correctly classified, had LR with the highest performance (0.88) compared to RF and SVC (both with 0.67), indicating that LR will correctly identify the occurrence of events 88% of the time. This is directly related to the fact that LR tended to generate more FP than FN compared to both RF and SVC. F1score, the harmonic mean between Recall and Precision, was very similar for the three models, which indicates that despite the differences, the models are comparable in terms of performance and behavior.

MCC, a coefficient to measure the quality of binary-classification problems, had SVC with the highest result (0.32), followed by RF (0.28) and LR (0.19). It can be noted that SVC and RF are performing with similar results, with SVC having an edge while LR is considerably behind.

From accuracy alone, it is not possible to find much difference between the models to determine which is best for this type of data. The OA (Overall Accuracy), which is the sum of Accuracy, F1score and MCC, is the measure to give a definitive result on the performance of the models. As was suggested by the trend in results, SVC was the best-performing model (1.65), followed closely by RF (1.58), and LR was further behind (1.47).

Overall, SVC provided the best-performing results all around, except for Recall, where LR was able to obtain a higher TP ratio, which is expected given the fact that LR had almost no FN. This advantage, however, was not enough to surpass the other methods and that's why SVC is the best method for analyzing this data.

However, the OA (Overall Accuracy) was not as high as there were Events with no corresponding precipitation, suggesting the Occurrence of Landslides on the island is not solely caused by Precipitation. It is important to note that the island is quite seismologically active, which could also cause landslides. For this reason, other variables like seismic activity and even soil moisture could be included as input to improve the performance of the models and evaluate their relationship with the occurrence of Events.

Our study conducted on Ometepe Island, Nicaragua, presents significant insights into the relationship between precipitation patterns and shallow landslide occurrences. The application of machine

learning models, namely Logistic Regression, Random Forest, and Support Vector Classifier, has facilitated a deeper understanding of this relationship. Notably, the Support Vector Classifier with a Sigmoidal kernel exhibited superior performance in correlating precipitation data with landslide events. This outcome underscores the potential of machine learning in enhancing landslide prediction and risk assessment, particularly in regions with complex geophysical characteristics like Ometepe Island.

A critical finding of our research is the correlation between precipitation and landslide events, particularly when considering a 7-day hydrometeorological period. This correlation is essential in understanding the triggering mechanisms of landslides in the region. However, our results also indicate that precipitation is not the only contributing factor to landslide occurrence, as seismic activity also plays a crucial role. This dual influence highlights the need for comprehensive risk assessment models that incorporate multiple environmental variables to accurately predict landslide occurrences.

Furthermore, our study contributes to the existing body of knowledge by emphasizing the importance of spatio-temporal aggregation of precipitation data in landslide analysis. The unique approach of combining SAR-derived coherence values with aggregated precipitation data provides a novel method for landslide prediction. This method could be particularly beneficial for regions with similar environmental and climatic conditions to Ometepe Island.

In light of our findings, future research should focus on integrating additional environmental variables, such as seismic data and soil moisture content, into landslide prediction models. This integration could enhance the accuracy and reliability of the models, thereby aiding in the development of effective landslide risk mitigation strategies. Additionally, extending this research to other regions with varying topographical and climatic conditions could provide further insights into the generalizability of our approach.

In conclusion, our study represents a significant step forward in the application of machine learning techniques for environmental hazard assessment. The methods and findings discussed here could serve as a foundation for future research aimed at improving landslide prediction and risk management strategies in vulnerable regions worldwide.

4. Conclusions

Landslides are one of the main hazards to society worldwide and Precipitation has been identified as the main cause for the majority of events. Thus, this study's main focus was to find the correlation between the Precipitation and the Shallow Landslide Ocurrence (SLO) in Ometepe Island, using spatio-temporal aggregation data and the Coherence Layer from InSAR images, as an alternative for interferograms with noise from vegetation.

On Phase 1, a database of Shallow Landslide Occurrence (SLO) from 2015 to 2020 for Ometepe Island was built. During this process, it was possible to see that the ideal threshold for coherence, later used to split data, is the one that gives the most even event ratio (close to 0.5), which should go in line with landslide data collected locally. On Phase 2, several coherence thresholds were evaluated to determine the one that performed best: in this case, 0.96. Finally on Phase 3, a deeper analysis was performed on this coherence threshold, and it was found that SVC (Support-Vector Classifier) was the best-performing model overall, followed closely by RF, with LR in last place. However, the results of this exploratory analysis and its application to the Ometepe Island show that, although precipitation is a strong force in the phenomenon, it is not possible to have conclusive results of identification with only this information.

In this study, we successfully developed a model that integrates precipitation data and remote sensing to classify shallow landslides in Ometepe Island. While the model demonstrated a correlation between precipitation and landslide events, it also revealed the complexity of predicting such phenomena based solely on precipitation data. One of the limitations we identified is the model's inability to account for other potential triggers of landslides, such as seismic activity and soil moisture

variations, which are prevalent in the region. This suggests a multidimensional approach for future studies, incorporating additional variables to enhance predictive accuracy.

Furthermore, we observed that certain landslide events occurred without corresponding high precipitation levels, indicating the influence of other factors. This finding underscores the need for a more holistic approach to studying shallow landslides, potentially integrating other environmental and geophysical data.

Our study makes a significant contribution to the field of shallow landslide research, particularly in the context of Ometepe Island. It opens up avenues for more comprehensive research that not only focuses on hydrometeorological aspects but also considers the broader geophysical characteristics of the region. Future research should aim to develop more robust models that can accurately forecast shallow landslides by considering a wider range of environmental factors. This is a contribution to the ongoing larger framework of Shallow Landslide studies and explores the possibility of using coherence as an indicator of the occurrence of a Shallow Landslide.

References

1. M.J., F.; D.N., P. Global Landslide Occurrence from 2004 to 2016 **2018**. pp. 2161–2181.
2. Haque, U., Blum, P., da Silva, P.F. et al.. Fatal landslides in Europe. *Landslides 13* **2016**, pp. 1545–1554.
3. Sassa, K. The Kyoto Landslide Commitment: First signatories. *Landslides* **2020**, *2019*, 2053–2057.
4. E. Van Eynde, S. Dondeyne, M.I.; Deckers, J.; Poesen, J. Impact of landslides on soil characteristics: Implications for estimating their age. *CATENA* **2017**, *157*, 173–179. doi:https://doi.org/10.1016/j.catena.2017.05.003.
5. Van Den Eeckhaut, M.; Hervás, J.; Montanarella, L., Landslide Databases in Europe: Analysis and Recommendations for Interoperability and Harmonisation. In *Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning*; Margottini, C.; Canuti, P.; Sassa, K., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp. 35–42. doi:10.1007/978-3-642-31325-7_4.
6. P., A.; R., C. Landslide hazard assessment: summary review and new perspectives. *Bull Eng Geol Env* **1999**, *58*, 21–44. doi:https://doi.org/10.1007/s100640050066.
7. IFAD. Investing in rural people in Nicaragua. https://www.ifad.org/documents/38714170/39972349/Investing+in+rural+people+in+Nicaragua_e.pdf/fcf638b4-8cbb-4c69-b0e1-589040114ce8, 2017.
8. Kreft, E..F. *Global Climate Risk Index 2016: Who suffers the Most from Weather Events? Weather-related Loss Events in 2014 and from 1995 to 2014.*; Berlin: German Watch, 2015.
9. Instituto de Geología y Geofísica (IGG). Proyecto de Reducción de Riesgo. Un Análisis de Susceptibilidad, Vulnerabilidad y Gobernanza de Riesgos en el Área de Influencia del Volcán Concepción, Nicaragua., 2017.
10. Chen, L.; Ge, X.; Yang, L.; Li, W.; Peng, L. An Improved Multi-Source Data-Driven Landslide Prediction Method Based on Spatio-Temporal Knowledge Graph. *Remote Sensing* **2023**, *15*. doi:10.3390/rs15082126.
11. Liu, X.; Peng, Y.; Lu, Z.; Li, W.; Yu, J.; Ge, D.; Xiang, W. Feature-Fusion Segmentation Network for Landslide Detection Using High-Resolution Remote Sensing Images and Digital Elevation Model Data. *IEEE Transactions on Geoscience and Remote Sensing* **2023**, *PP*, 1–1. doi:10.1109/TGRS.2022.3233637.
12. Xia, J.; Liu, H.; Zhu, L. Landslide Hazard Identification Based on Deep Learning and Sentinel-2 Remote Sensing Imagery. *Journal of Physics: Conference Series* **2022**, *2258*, 012031. doi:10.1088/1742-6596/2258/1/012031.
13. Park, S.; Kim, J. Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Applied Sciences* **2019**, *9*, 942. doi:10.3390/app9050942.
14. Troncone, A.; Pugliese, L.; Conte, E. Rainfall Threshold for Shallow Landslide Triggering Due to Rising Water Table. *Water* **2022**, *14*. doi:10.3390/w14192966.
15. Nhu, V.H.; Mohammadi, A.; Shahabi, H.; Ahmad, B.B.; Al-Ansari, N.; Shirzadi, A.; Geertsema, M.; R. Kress, V.; Karimzadeh, S.; Valizadeh Kamran, K.; Chen, W.; Nguyen, H. Landslide Detection and Susceptibility Modeling on Cameron Highlands (Malaysia): A Comparison between Random Forest, Logistic Regression and Logistic Model Tree Algorithms. *Forests* **2020**, *11*. doi:10.3390/f11080830.
16. Adnan, M.S.G.; Rahman, S.; Ahmed, N.; Ahmed, B.; Rabbi, M.F.; Rahman, R.M. Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sensing* **2020**, *12*, 3347. doi:10.3390/rs12203347.
17. Nhu, V.H.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Jaafari, A.; Chen, W.; Miraki, S.; Dou, J.; Luu, C.; Górski, K.; Thai Pham, B.; Nguyen, H.D.; Ahmad, B.B. Shallow Landslide Susceptibility

Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naïve Bayes Tree, Artificial Neural Network, and Support Vector Machine Algorithms. *International Journal of Environmental Research and Public Health* **2020**, *17*. doi:10.3390/ijerph17082749.

18. Ma, Z.; Mei, G.; Piccialli, F. Machine learning for landslides prevention: a survey. *Neural Computing and Applications* **2020**, *33*, 10881–10907. doi:10.1007/s00521-020-05529-8.

19. Nocentini, N.; Rosi, A.; Segoni, S.; Fanti, R. Towards landslide space-time forecasting through machine learning: the influence of rainfall parameters and model setting. *Frontiers in Earth Science* **2023**, *11*. doi:10.3389/feart.2023.1152130.

20. Luo, X.; Lin, F.; Chen, Y.; Zhu, S.; Xu, Z.; Huo, Z.; Yu, M.; Peng, J. Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features. *Scientific Reports* **2019**, *9*. doi:10.1038/s41598-019-51941-z.

21. Ge, P.; Gokon, H.; Meguro, K.; Koshimura, S. Study on the intensity and coherence information of high-resolution alos-2 sar images for rapid massive landslide mapping at a pixel level. *Remote Sensing* **2019**, *11*, 2808. doi:10.3390/rs11232808.

22. Shih, P.; Wang, H.; Li, K.W.; Liao, J.N.; Pan, Y. Landslide monitoring with interferometric sar in liugui, a vegetated area. *Terrestrial, Atmospheric and Oceanic Sciences* **2019**, *30*, 521–530. doi:10.3319/tao.2019.04.13.01.

23. Azarafza, M.; Azarafza, M.; Akgün, H.; Atkinson, P.M.; Derakhshani, R. Deep learning-based landslide susceptibility mapping. *Scientific reports* **2021**, *11*. doi:10.1038/s41598-021-03585-1.

24. A., S.; M.A., U.; E., P. *Rasgos Territoriales del turismo de la isla de Ometepe, Nicaragua*; Universidad de Murcia, 2008; pp. 159 – 179.

25. UNESCO. Retrieval from Ometepe Island Biosphere Reserve, Nicaragua. https://en.unesco.org/biosphere/lac/ometepe-island, 2021.

26. C., A.; E., A. Suelos, Capacidad de Uso de la Tierra y Conflictos de Uso en el Municipio Moyogalpa. https://cenida.una.edu.ni/relectronicos/RENP30A284s.pdf, 2009.

27. Devoli, G.; Strauch, W.; Chávez, G.; Hoeg, K. A landslide database for Nicaragua: a tool for landslide-hazard management. *Landslides* **2007**, *4*. doi:https://doi.org/10.1007/s10346-006-0074-8.

28. C., F.; P., P.; et al., L.M. The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes **2015**. doi:https://doi.org/10.1038/sdata.2015.66.

29. Sharma, A.; Sharma, K.K.; Sapate, S.G. A prototype model for detection and classification of landslides using satellite data. *Journal of Physics: Conference Series* **2022**, *2327*, 012029. doi:10.1088/1742-6596/2327/1/012029.

30. Park, S.; Lee, S. On the use of single-, dual-, and quad-polarimetric sar observation for landslide detection. *ISPRS International Journal of Geo-Information* **2019**, *8*, 384. doi:10.3390/ijgi8090384.

31. Ray, R.L.; Lazzari, M.; Olutimehin, T. Remote sensing approaches and related techniques to map and study landslides. *Landslides - Investigation and Monitoring* **2020**. doi:10.5772/intechopen.93681.

32. Xia, J.; Liu, H.; Zhu, L. Landslide hazard identification based on deep learning and sentinel-2 remote sensing imagery. *Journal of Physics: Conference Series* **2022**, *2258*, 012031. doi:10.1088/1742-6596/2258/1/012031.

33. Lv, P.; Ma, L.; Li, Q.; Du, F. Shapeformer: a shape-enhanced vision transformer model for optical remote sensing image landslide detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**, *16*, 2681–2689. doi:10.1109/jstars.2023.3253769.

34. Mondini, A.C.; Santangelo, M.; Rocchetti, M.; Rossetto, E.; Manconi, A.; Monserrat, O. Sentinel-1 sar amplitude imagery for rapid landslide detection. *Remote Sensing* **2019**, *11*, 760. doi:10.3390/rs11070760.

35. Kavats, O.; Khramov, D.; Sergieieva, K.; Vasyliev, V.V. Monitoring of sugarcane harvest in brazil based on optical and sar data. *Remote Sensing* **2020**, *12*, 4080. doi:10.3390/rs12244080.

36. Amitrano, D.; Guida, R.; Martino, G.D.; Iodice, A. Glacier monitoring using frequency domain offset tracking applied to sentinel-1 images: a product performance comparison. *Remote Sensing* **2019**, *11*, 1322. doi:10.3390/rs11111322.

37. Gama, F.F.; Cantone, A.; Mura, J.C. Monitoring horizontal and vertical components of samarco mine dikes deformations by dinsar-sbas using terrasar-x and sentinel-1 data. *Mining* **2022**, *2*, 725–745. doi:10.3390/mining2040040.

38. Luca, C.D.; Casu, F.; Manunta, M.; Onorato, G.; Lanari, R. Comments on "study of systematic bias in measuring surface deformation with sar interferometry". *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–5. doi:10.1109/tgrs.2021.3103037.

39. Tarigan, S.; Kristanto, Y.; Utomo, W.Y. Characterizing subsidence in used and restored peatland with sentinel sar data. *Frontiers in Environmental Science* **2023**, *11*. doi:10.3389/fenvs.2023.1088923.

40. Haghighi, M.H.; Motagh, M. Assessment of ground surface displacement in taihape landslide, new zealand, with c- and x-band sar interferometry. *New Zealand Journal of Geology and Geophysics* **2016**, *59*, 136–146. doi:10.1080/00288306.2015.1127824.

41. Dou, C.; Guo, H.; Chen, R.; Yue, X.; Zhao, Y. Integrated data processing methodology for airborne repeat-pass differential sar interferometry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2014**, *XL-1*, 109–112. doi:10.5194/isprsarchives-xl-1-109-2014.

42. Gebremichael, E.; Hernandez, R.; Alsleben, H.; Ahmed, M.; Denne, R.; Harvey, O. Kinematics and controlling factors of slow-moving landslides in central texas: a multisource data fusion approach. *Geosciences* **2024**, *14*, 133. doi:10.3390/geosciences14050133.

43. Tong, X.; Sandwell, D.T.; Fialko, Y. Coseismic slip model of the 2008 wenchuan earthquake derived from joint inversion of interferometric synthetic aperture radar, gps, and field data. *Journal of Geophysical Research: Solid Earth* **2010**, *115*. doi:10.1029/2009jb006625.

44. Corsa, B.D.; Barba-Sevilla, M.; Tiampo, K.F.; Meertens, C.M. Integration of dinsar time series and gnss data for continuous volcanic deformation monitoring and eruption early warning applications. *Remote Sensing* **2022**, *14*, 784. doi:10.3390/rs14030784.

45. Scherer, N.F.; Carlson, R.J.; Matro, A.; Du, M.; Ruggiero, A.J.; Romero-Rochín, V.; Cina, J.A.; Fleming, G.R.; Rice, S.A. Fluorescence-detected wave packet interferometry: time resolved molecular spectroscopy with sequences of femtosecond phase-locked pulses. *The Journal of Chemical Physics* **1991**, *95*, 1487–1511. doi:10.1063/1.461064.

46. De, A.K.; Roy, D.; Goswami, D. Fluorophore discrimination by tracing quantum interference in fluorescence microscopy. *Physical Review A* **2011**, *83*. doi:10.1103/physreva.83.015402.

47. Bazzi, H.; Baghdadi, N.; Najem, S.; Jaafar, H.; Page, M.L.; Zribi, M.; Faraslis, I.; Spiliotopoulos, M. Detecting irrigation events over semi-arid and temperate climatic areas using sentinel-1 data: case of several summer crops. *Agronomy* **2022**, *12*, 2725. doi:10.3390/agronomy12112725.

48. Raspini, F.; Ciampalini, A.; Conte, S.D.; Lombardi, L.; Nocentini, M.; Gigli, G.; Ferretti, A.; Casagli, N. Exploitation of amplitude and phase of satellite sar images for landslide mapping: the case of montescaglioso (south italy). *Remote Sensing* **2015**, *7*, 14576–14596. doi:10.3390/rs71114576.

49. Tzouvaras, M.; Danezis, C.; Hadjimitsis, D. Small scale landslide detection using sentinel-1 interferometric sar coherence. *Remote Sensing* **2020**, *12*, 1560. doi:10.3390/rs12101560.

50. Sevgen, E.; Kocaman, S.; Nefeslioğlu, H.A.; Gökçeoğlu, C. A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ann and random forest. *Sensors* **2019**, *19*, 3940. doi:10.3390/s19183940.

51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

52. Sean Gillies et al. Rasterio: geospatial raster I/O for Python programmers. https://github.com/mapbox/rasterio, 2013.

53. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M.H.; Brett, M.; Haldane, A.; del R'ıo, J.F.; Wiebe, M.; Peterson, P.; G'erard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T.E. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. doi:10.1038/s41586-020-2649-2.

54. Silversmith, W. cc3d: Connected Components on Multilabel 3D Images. Taken from: https://github.com/seung-lab/connected-components-3d/, 2021.

55. Wang, H.; Zhang, L.; Yin, K.; Luo, H.; Li, J. Landslide identification using machine learning. *Geoscience Frontiers* **2021**, *12*, 351–364. doi:https://doi.org/10.1016/j.gsf.2020.02.012.