DOI: https://doi.org/10.5592/CO/EUROENGEO.2024.211

MACHINE LEARNING LANDSLIDE SUSCEPTIBILITY MAPPING IN WESTERN GREECE WITH INSAR TIME-SERIES ANALYSIS

STAVROULA ALATZA[1,2], CONSTANTINOS LOUPASAKIS[1], ALEXIS APOSTOLAKIS[2,3], CHARALAMPOS KONTOES[2], MARTHA KOKKALIDOU[2], NIKOLAOS S. BARTSOTAS[2], CONSTANTINOS NEFROS[1]

[1] *National Technical University of Athens (NTUA), School of Mining and Metallurgical Engineering, Department of Geological Sciences, Greece,* kostasnefros@central.ntua.gr, cloupasakis@metal.ntua.gr

[2] *National Observatory of Athens, Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, Operational Unit "BEYOND Center of Earth Observation Research and Satellite Remote Sensing", Greece,* alatza@noa.gr, alex.apostolakis@noa.gr, kontoes@noa.gr, m.kokkalidou@noa.gr, nbartsotas@noa.gr

[3] *National Technical University of Athens, School of Electrical and Computer Engineering*

Abstract

The current study aims to investigate landslide susceptibility in Western and Central Greece and to enrich the existing knowledge in slow-moving deformation phenomena, mainly triggered by prolonged or excess precipitation events. For this purpose, SAR interferometry and Artificial Intelligence are implemented on Earth Observation data. Landslide susceptibility mapping was performed in a broad area within the most landslide prone geotectonic zones of Greece, extending from Crete Island up to the Greek - Albanian borders. These are the Ionios, Gavrovo and Pindos geotectonic zones. The variety of geological formations and morphological features, crucial infrastructure and the plethora of settlements provide a great variety of conditions and an exceptionally large landslide inventory. By exploiting a national scale inventory of Line-of-Sight surface displacements in Greece, the so-called InSAR Greece project, more than 3000 landslides were detected in Western Greece. Topographical, geological, meteorological, hydrological parameters and vegetation, were introduced to the model, as landslide contributing factors. The prediction problem was approached as a binary classification problem, whereby two classes were defined: (1) landslide and (2) no landslide for each assessed data point. The landslide prediction was solved by training robust classic machine learning algorithms using the aforementioned dataset. The annotation of the dataset was performed by experts in the field. A strict Machine Learning methodology was employed for training, tuning and testing the machine learning, comprising of 5-fold cross validation, hyperparameter tuning, feature ranking and selection and strict train validation dataset split to avoid spatial autocorrelation. XGBoost algorithm showed the best performance among the traditional ML algorithms, thus it was preferred for producing susceptibility maps. The exploitation of big volumes of EO data and AI, in landslide susceptibility mapping, can provide a valuable tool in risk reduction. Also, validation of susceptibility mapping with ground truth investigations, provides an additional advantage in establishing a landslide susceptibility system and in the adoption of mitigation measures.

Key words

XGBoost, landslide susceptibility, InSAR, machine learning.

1 Introduction

InSAR time-series analysis, has long been identified as an effective tool for landslides (LS) detection and monitoring (Aslan et al., 2020; Bekaert et al., 2020; Nefros et al., 2023; Kontoes et al., 2021; Smail et al., 2022). On the other hand, AI has rapidly demonstrated its efficacy in predicting LS -prone zones and providing early warning systems for LS. Numerous AI algorithms, predominantly tree-based ensemble models, have been utilized in the field of LS susceptibility mapping. Support vector machines (SVM) (Lee et al., 2017), Random Forest (RF) models (Park et al., 2019), boost extreme gradient boosting (XGBoost) models (Can et al., 2021), are among the most commonly used ML models for LS detection and susceptibility mapping. Deep learning models have also demonstrated their efficiency in LS susceptibility studies (Zhang et al., 2024; Habumugisha et al.,2022).

Greece is listed among the most landslide-prone countries in the Mediterranean, due to its intense geomorphology. LS are considered among the most destructive natural hazards worldwide, often associated with seismic events or volcanic eruptions. In recent decades, extreme precipitation events intensified by the climate change, urban activities, and deforestation, along with seismic activity or volcanic eruptions, are considered as significant contributing factors for LS. In Greece Pindos, Gavrovo and Ionian are three of the most LS prone geotectonic zones. In Pindos geotectonic zone more than 40% of the total LS in Greece are recorded. On the contrary, in Gavrovo and Ionian zones, LS occurrences of 4% and 4.5%, are recored (Koukis et al., 2005; Sabatakakis et al., 2013). Intense tectonism, present at Pindos geotectonic zone (Delibasis et al., 1977), steep slopes, the highly susceptible flysch, in combination with high precipitation levels, enhance the frequency of slope failures. On the other hand, the smooth tectonic structures (Mountrakis, 2010) in the Gavrovo and Ionian geotectonic zones and the existence of limestones decrease the occurrence of LS in these zones. In the present study, the LS inventory is formed with landslides detected in Aitoloakarnania and Evritania prefectures, in Western and Central Greece (Figure 1), using InSAR time-series analysis and satellite image investigations from geotechnical experts. A strict ML methodology was employed for LS susceptibility mapping.

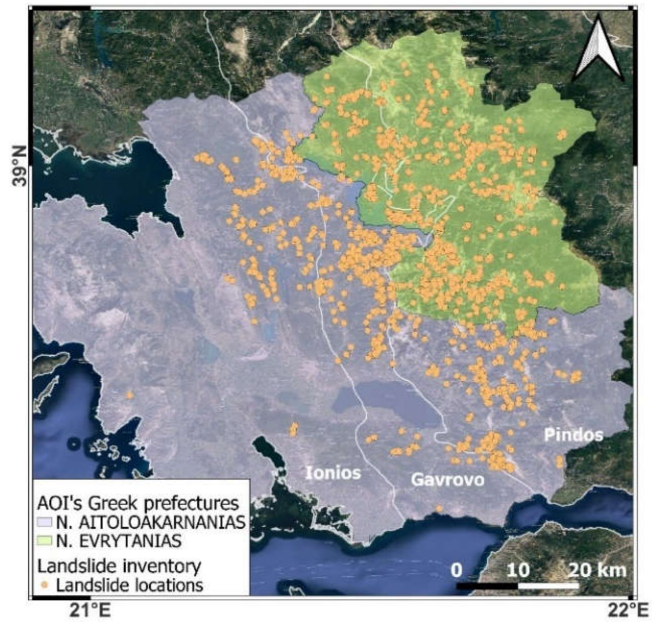

Figure 1. The study area and the landslides inventory.

As in most machine learning problems, the performance of ML models in landslide susceptibility studies heavily relies on the volume and accuracy of the dataset. An additional challenging aspect in ML studies, relates to the migration of the model to a larger scale. The present study exploits an exceptionally large LS inventory, with more than 3000 validated landslides created for LS susceptibility mapping in Western and Central Greece, using the XGBoost algorithm. The geotectonic zone index is added to the model as a contributing factor. By adapting the model to the special characteristics of each geotectonic zone, the migration of the model to a larger scale, will be facilitated, in future work. Finally, the exploitation of time-series InSAR techniques served as a valuable tool, in the creation of the LS inventory and also as a validation for the model's predictions, highlighting the significant potential of integrating remote sensing techniques and AI.

2 Dataset

2.1 LS locations

LS locations were derived from multiple sources, including landslides from Kontoes et al., (2021), enriched by 704 more landslides identified by satellite image investigations by geotechnical experts. 2146 slow-moving LS were derived by Line-of-Sight (LOS) displacements from the InSAR Greece project (Papoutsis et al., 2020). In InSAR Greece, LOS displacements were estimated for Greece's mainland, using Sentinel-1 SLCs from 2015 to 2019, with a parallelized version of the Stanford method for Persistent Scatterers Interferometry (Hooper et al., 2004). All LS locations introduced in the ML model, were finally validated by geotechnical expert. A total of 3224 LS forms an exceptionally large LS inventory in Aitoloakarnania and Evritania prefectures. It is worthwhile noting that one limitation of the PSI technique employed to identify LS locations, is the low presence of permanent scatterers in vegetated areas. Therefore, in the present study, LS locations derived from InSAR techniques are not detected in areas with dense vegetation.

The no-LS locations, were also derived from the Greece InSAR product. Permanent Scatterers with almost zero LOS displacement rates, were identified. To avoid misinterpretation of PS points as LS and no LS locations and to avoid data leakage because of spatial autocorrelation (Roberts et al., 2016) between close PS points during the training, buffers of several sizes, ranging from 200m to 1km, were tested around the LS and no LS locations. A buffer of 600m proved to be effective in avoiding labelling multiple PS points located in the same sliding slope, as different LS locations. Similarly, a buffer of 400m around each no LS PS point was applied to ensure that there is no spatial correlation in the final LS susceptibility map product. Therefore, 2753 no LS locations were introduced to the ML model. In the selection of non-LS points, two critical aspects were considered. First, the number of LS and non-LS points to have a ratio of almost 1:1. Second, the spatial distribution of LS and no LS locations to be homogeneous.

2.2 LS contributing factors

LS occurrence is triggered by both natural and anthropogenic factors. In Greece, geology (Sikodimou et al., 2018), earthquakes (Papathanassiou et al., 2021) and rainfall (Nefros et al., 2023), are considered significant LS contributing factors. In the present study, the LS contributing factors introduced to the model are summarized in Table 1. The LS contributing factors were selected based on the special characteristics of the study area (intense geomorphology, high precipitation rates, etc.) and literature review. These include, geomorphological and geological factors, hydrological and topography factors, climatological factors, land use/land cover and vegetation. The EU-DEM, with a 25m cell size, was employed for the estimation of the elevation factors. The same 25m pixel size was maintained in all computed raster layers. Among the geological factors, LS factor (Slope Length and Steepness factor), that is associated with soil erosion and surface lithology was also added. Land cover, as an indicator of natural or manmade elements in the study area, is also an important LS contributing factor. NDVI, as an indication of existence or absence of vegetation, was used as an LS contributing factor in the ML LS susceptibility model. Regarding the climate factors, precipitation is considered as a critical factor as the majority of the recorded LS in the study area, are triggered by extreme precipitation events. Snowmelt also poses a serious threat in slope stability. Hydrological factors which relate to soil moisture, surface roughness, slope erosion from water flow and sediment transportation, are also added as important LS contributing factors.

Table 1. LS contributing factors

Contributing Factors	Source
Elevation Aspect Slope	EU-DEM (25m)
LS factor (Slope Length and Steepness factor)	EUROPEAN SOIL DATA CENTRE (ESDAC)
Land use - Land cover	Copernicus Corine Land Cover 2018
Surface Lithology	EGDI 1:1 Million pan-european Surface Geology INSPIRE conformant National WFS services on GeologicUnit
Snow melt	ERA5 – LAND (spatial sampling of 9 km)
Precipitation	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) from 2000 to 2020 (spatial sampling of 5 km)
Sediment transport index (STI) Topographic wetness index (TWI) Terrain Ruggedness Index (TRI) Stream power index (SPI)	Calculated from the DEM (25m)
Normalized difference vegetation index (NDVI)	MODIS data (1km)

3 Methods

Among the traditional machine learning algorithms tested to evaluate the dataset's potential, the XGBoost algorithm was preferred due to its superior performance in the study area. Another reason was the advantage of XGBoost compared to algorithms, like AdaBoost and Gradient Boosting, to prevent overfitting, owing to regularization techniques (Lasso-1, Lasso-2). Feature selection is an important part of a landslide classification problem, as uncorrelated factors may introduce noise to the model and negatively affect its performance. The selection of the final contributing factors, was determined after model training and feature importance estimation. Considering the different scales of the input features, a normalization of the dataset was applied, using a standard range from -1 to +1. To simplify data complexity, particularly for ensemble tree methods, and to ensure the dataset’s compatibility with a wide range of ML algorithms, discretization followed by One Hot Encoding was applied in aspect, slope, LS factor, land use/land cover and geology factors. A 70% of LS and no LS locations was used as a train dataset and 30% was used as a test dataset for validation. To better support the generalization of the model a 5-fold cross validation was applied on the training dataset along with model hyperparameterization. The top performing model from cross validation was used for prediction on the test dataset. The final dataset and the split ratio are provided in Figure 2.

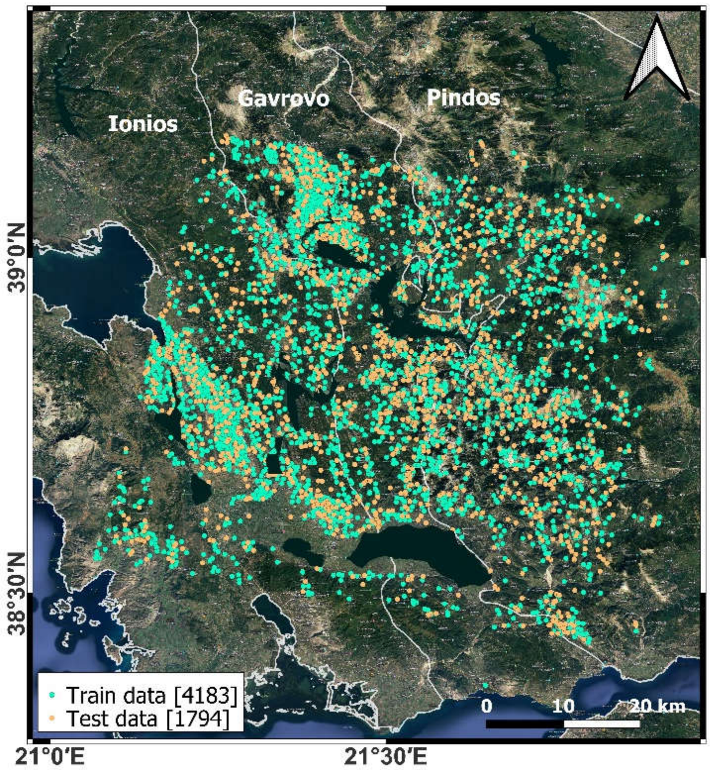

Figure 2. Dataset split ratio.

4 Results

Model evaluation is a critical aspect in every ML problem, especially in geological hazards applications, to ensure the reliability of the model's predictions. The XGBoost model used for LS predictions in the study area, was mainly evaluated based on precision, recall, F1 and Mathew's correlation coefficient (MCC) metrics. The MCC, is considered an important metric for the model's performance, owing to its ability to provide a high score, taking into account good results in all four components (Chicco et al., 2020) of the confusion matrix (Figure 3). Table 2 presents the model's evaluation metrics on the test dataset. The high MCC value of 0.75 in combination with the other metrics, verify the good performance of the model in landslide susceptibility predictions.

Table 2. XGBoost evaluation metric results on the test dataset

Score metric	Precision	Recall	F1	Support	Accuracy	MCC
No Landslide	0.94	0.78	0.85	827	0.88	0.75
Landslide	0.83	0.96	0.89	967		

According to the precision results shown in Table 2, 83% of the predictions were true positives and based on the recall, 96% of the identified landslides were actual events. The high F1 score verifies the ability of the model to effectively identify actual LS events and minimize false positive events. A confusion matrix was also used to assess the model's performance (Figure 3). Based on Figure 3, the model performed well in identifying True Negative events, which are no landslides and True Positive events that correspond to actual landslide events, in the test dataset.

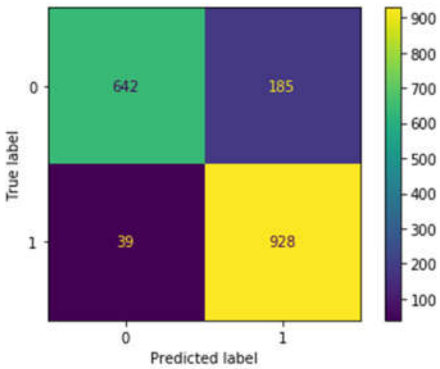

Figure 3. Confusion matrix

The final landslide susceptibility map, based on predictions on a grid with a cell size of 25m is presented in Figure 4. The landslide locations of the validation dataset are also added, as an indicator of the model's performance. As Figure 4 illustrates, the majority of landslides are presented in Gavrovo and Pindos geotectonic zones, where landslide occurrence is more frequent, due to the geotectonic evolution of the zones and the geotechnical characteristics of the geological formations.

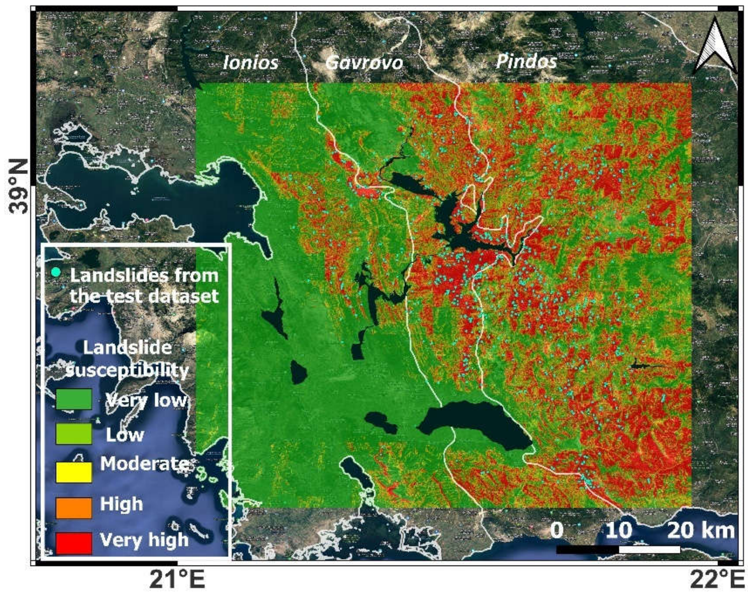

Figure 4. LS susceptibility map

5 Conclusions

This was a first attempt to demonstrate the potential of InSAR data, AI and ground truth investigations in establishing an LS susceptibility system in Greece. Using an originally created dataset LS susceptibility mapping with ML methodology was performed in a broad area in Western and Central Greece, within the most LS prone geotectonic zones, that is Ionios, Gavrovo and Pindos zones. Landslide locations from InSAR timeseries analysis in Greece and ground truth inspections formed an LS inventory of more than 3000 LS. Topographical, geological, meteorological, hydrological parameters and vegetation, were introduced to the LS susceptibility model, as landslide contributing factors. The XGBoost algorithm was used to perform LS susceptibility predictions and produce the LS susceptibility maps. The model demonstrates high performance with an accuracy of 88%, a precision of 83% in predicting landslide occurrences, and F1-scores of 0.85 for non-landslide events and 0.89 for landslide events. Also, the addition of the geotectonic zone index as a contributing parameter, is an important factor to facilitate the model's predictions in new areas in Western and Central Greece, which will be included in future work.

Acknowledgements
The authors acknowledge the ESA Contract_ SENTINELS COLLABORATIVE GS DATA ACCESS ESA NODE, No. 4000142206/23/I-NS running by Operational Unit BEYOND/IAASARS/NOA and the "Support for upgrading the operation of the National Network for Climate Change (Climpact)". Climpact is financed by the National section of the PDE National Development Program 2021-2025 Ministry of Development – General Secretariat of Research and Innovation.

References

Aslan, G.; Foumelis, M.; Raucoules, D.; De Michele, M.; Bernardie, S.; Cakir, Z. Landslide Inventory Mapping and Monitoring Using Persistent Scatterer Interferometry (PSI) Technique in the French Alps. *Remote Sensing*. 2020, 12(8), 1305, https://doi.org/10.3390/rs12081305. IF: 4.118.

Ballabio, C.; Sterlacchini, S. Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. *Mathematical geosciences*. 2012, 44. 10.1007/s11004-011-9379-9.

Bekaert, D.P.S.; Handwerger, A. L.; Agram, P.; Kirschbaum, D. B. InSAR-based detection method for mapping and monitoring slow-moving landslides in remote regions with steep and mountainous terrain: An application to Nepal, *Remote Sensing of Environment.* 2020, 249 https://doi.org/10.1016/j.rse.2020.111983.

Can, R.; Kocaman, S.; Gokceoglu, C. A Comprehensive Assessment of XGBoost Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam, Turkey. *Appl. Sci*. 2021, 11, 4993. https://doi.org/10.3390/app11114993.

Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020, 21, 6. https://doi.org/10.1186/s12864-019-6413-7

Delibasis, N.; Karydis, P. Recent earthquake activity in Trichonis region and its tectonic significance. *Ann. Geofis*. 1977, 30, 19 – 81.

Habumugisha, J.M.; Chen, N.; Rahman, M.; Islam, M.M.; Ahmad, H.; Elbeltagi, A.; Sharma, G.; Liza, S.N.; Dewan, A. Landslide Susceptibility Mapping with Deep Learning Algorithms. *Sustainability* 2022, 14, 1734. https://doi.org/10.3390/su14031734.

Hooper, A.; Zebker, H.; Segall, P.; Kampes, B. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophys. Res. Lett*. 2004, 31, L23611. https://doi.org/10.1029/2004GL021737.

Kontoes, C.; Loupasakis, C.; Papoutsis, I.; Alatza, S.; Poyiadji, E.; Ganas, A.; Psychogyiou, C.; Kaskara, M.; Antoniadi, S.; Spanou, N. Landslide Susceptibility Mapping of Central and Western Greece, Combining NGI and WoE Methods, with Re-mote Sensing and Ground Truth Data. *Land*. 2021, 10, 402. https://doi.org/10.3390/land10040402.

Koukis, G.; Sabatakakis, N.; Nikolaou, N., Loupasakis, C. Landslide hazard zonation in Greece. *In Proceedings of open symposium on landslide risk analysis and sustainable disaster management in the First General Assembly of International Consortium on Landslides*. 2005, 291-296.

Lee, S.; Hong, S.-M.; Jung, H.-S. A Support Vector Machine for Landslide Susceptibility Mapping in Gangwon Province, Ko-rea. *Sustainability.* 2017, 9, 48. https://doi.org/10.3390/su9010048.

Mountrakis, D. Γεωλογια και Γεωτεκτονικη Εξελιξη της Ελλαδας. UNIVERSITY STUDIO PRESS, Thessaloniki, 2010, 374.

Nefros, C.; Tsagkas, D.S.; Kitsara, G.; Loupasakis, C.; Giannakopoulos, C. Landslide Susceptibility Mapping under the Climate Change Impact in the Chania Regional Unit, West Crete, Greece. *Land* 2023, 12, 154. https://doi.org/10.3390/land12010154

Nefros, C.; Alatza, S.; Loupasakis, C.; Kontoes, C. Persistent Scatterer Interferometry (PSI) Technique for the Identification and Monitoring of Critical Landslide Areas in a Regional and Mountainous Road Network. *Remote Sensing.* 2023, 15, 1550. https://doi.org/10.3390/rs15061550.

Papathanassiou, G.; Valkaniotis, S.; Ganas, A. Spatial patterns, controlling factors, and characteristics of landslides triggered by strike-slip faulting earthquakes: case study of Lefkada island, Greece. *Bull Eng Geol Environ*. 2021, 80, 3747–3765. https://doi.org/10.1007/s10064-021-02181-x

Papoutsis, I.; Kontoes, C.; Alatza, S.; Apostolakis, A.; Loupasakis, C. InSAR Greece with Parallelized Persistent Scatterer In-terferometry: A National Ground Motion Service for Big Copernicus Sentinel-1 Data. *Remote Sensing*. 2020, 12, 3207. https://doi.org/10.3390/rs12193207

Park, S.; Kim, J. Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of Their Performance. Appl. Sci. 2019, 9, 942. https://doi.org/10.3390/app9050942.

Roberts, D; Bahn, V.; Ciuti, S.; Boyce, M.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.; Schröder, B.; Thuiller, W.; Warton, D.; Wintle, B.; Hartig, F.; Dormann, C. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*. 2016, 40. https://doi.org/10.1111/ecog.02881.

Sabatakakis, N.; Koukis, G.; Vassiliades, E.; Lainas, S. Landslide susceptibility zonation in Greece. *Nat. Hazards*. 2013, 65, 523–543.

Skilodimou, H.D.; Bathrellos, G.D.; Koskeridou, E.; Soukis, K.; Rozos, D. Physical and Anthropogenic Factors Related to Landslide Activity in the Northern Peloponnese, Greece. *Land*. 2018, 7, 85. https://doi.org/10.3390/land7030085.

Smail, T.; Abed, M.; Mebarki, A.; Lazecky, M. Earthquake-induced landslide monitoring and survey by means of InSAR, *Nat. Hazards Earth Syst. Sci.* 2022, 22, 1609–1625, https://doi.org/10.5194/nhess-22-1609-2022.

Zhang, Q.; Wang, T. Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities. *Remote Sens*. 2024, 16, 1344. https://doi.org/10.3390/rs16081344.