environmental sciences proceedings

MDPI

Proceeding Paper

Automatic Classification of Active Deformation Areas Based on Synthetic Aperture Radar Data and Environmental Covariates Using Machine Learning—Application in SE Spain †

Jhonatan Rivera-Rivera [1,2,*], Marta Béjar-Pizarro [2], Héctor Aguilera [2], Carolina Guardiola-Albert [2], César Husillos [2], Pablo Ezquerro [2], Anna Barra [3], Rosa María Mateos [2], María Cuevas-González [3], Roberto Sarro [2], Oriol Monserrat [3], Mónica Martínez-Corbella [2], Michele Crosetto [3] and Juan López-Vinielles [2]

1 ETSI Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid (UPM), 28031 Madrid, Spain
2 Geological Survey and Mining Institute of Spain (CN IGME-CSIC), Rios Rosas 23, 28003 Madrid, Spain; m.bejar@igme.es (M.B.-P.); h.aguilera@igme.es (H.A.); c.guardiola@igme.es (C.G.-A.); c.husillos@igme.es (C.H.); p.ezquerro@igme.es (P.E.); rm.mateos@igme.es (R.M.M.); r.sarro@igme.es (R.S.); m.martinez@igme.es (M.M.-C.); j.lopez@igme.es (J.L.-V.)
3 Geomatics Research Unit, Centre Tecnológic Telecomunicacions Catalunya (CTTC/CERCA), 08860 Castelldefels, Spain; abarra@cttc.cat (A.B.); maria.cuevas@cttc.cat (M.C.-G.); omonserrat@cttc.cat (O.M.); michele.crosetto@cttc.cat (M.C.)
* Correspondence: js.rivera@igme.es; Tel.: +34-603-826-324

† Presented at the IV Conference on Geomatics Engineering, Madrid, Spain, 6–7 July 2023.

Abstract: Deformation processes, both natural (e.g., subsidence, landslides, active tectonics) and induced (e.g., associated with mining, construction. groundwater exploitation), result in significant socioeconomic losses worldwide. Accurate detection and classification of these processes are crucial for effective risk management. In this study, we present a novel approach for the automatic classification of deformation processes using Interferometric Synthetic Aperture Radar (InSAR) data and machine learning techniques. Specifically, we use a decision tree-based classification algorithm to train a model capable of recognizing and distinguishing different types of deformation processes using time series of displacements, grouped into Active Deformation Areas (ADAs). We test this methodology in a large area in SE Spain. Our results demonstrate promising performance, with an Area Under the Curve (AUC) > 0.95, identifying several covariates of morphometric, geological, hydrogeological, and geotechnical nature as key factors. This automatic classification of InSAR data holds significant implications for risk management associated with ground deformation, providing a potentially valuable tool for decision makers in urban planning and land management officials.

Keywords: geohazards; Displacement Time Series; SAR; ADAs; Time Series Clustering; machine learning; SE Spain

check for updates

Citation: Rivera-Rivera, J.; Béjar-Pizarro, M.; Aguilera, H.; Guardiola-Albert, C.; Husillos, C.; Ezquerro, P.; Barra, A.; Mateos, R.M.; Cuevas-González, M.; Sarro, R.; et al. Automatic Classification of Active Deformation Areas Based on Synthetic Aperture Radar Data and Environmental Covariates Using Machine Learning—Application in SE Spain. *Environ. Sci. Proc.* **2023**, *28*, 15. https://doi.org/10.3390/environsciproc2023028015

Academic Editors: María Belén Benito Oterino, José Fernández Torres, Rosa María García Blanco, Jorge Miguel Gaspar Escribano, Miguel Ángel Manso Callejo and Antonio Vázquez Hoehne

Published: 29 December 2023

1. Introduction

The detection and classification of active deformation areas is a novel approach that allows non-expert users of InSAR to integrate SAR-based products into risk management. Bonì et al. [1] and Barra et al. [2] established the initial methodologies for the automatic detection of Active Deformation Areas (ADAs) using GIS tools. Bonì et al. [3] implemented their methodology using ArcGIS, while Navarro et al. [4] implemented Barra's methodology in a software package with a graphical user interface called ADAfinder (V2.0.9 is the last version and it's available free on request), using the C++ programming language. ADAfinder determines active Deformational Time Series (DTS) through standard deviation thresholds, isolation distance, and average velocity. Subsequently, it groups them into polygonal clusters (ADAs), whose dimensions depend on parameters such as the defined influence radius and the minimum number of DTS required to form an ADA. Additionally, ADAfinder calculates a quality index for each ADA.

Tomás et al. [5] developed ADAclassifier (V2.0.9 is the last version and it's available free on request), a software package that determines whether the deformation of an ADA is related, potentially related, or unrelated to a sliding, sinkhole, subsidence, or settlement process. The classification is determined using a heuristic decision tree based on intersection thresholds with inventories of processes (landslides, subsidence, and sinkholes), infrastructure, and geological variables (Quaternary deposits and saline-carbonate soils/rocks), as well as thresholds for the horizontal velocity, slope, and coefficient-of-fit correlation of the DTS to a negative exponential function.

Recently, Festa et al. [6] proposed a machine learning-based methodology to classify DTS (instead of ADAs) into three processes: subsidence, landslide, and deformation related to mining. In this methodology, random forest is trained with morphometric variables (slope, aspect, elevation, Topographic Wetness Index (TWI), profile curvature, general curvature, and plan curvature), variables related to inventories (distance to landslides and mining sites), a geological variable (lithology), and a variable that describes the ratio between horizontal (E-W) and vertical velocity, called KVH, useful for distinguishing landslides from subsidence.

In this study, we combined the inherent advantages of each approach to achieve the automatic classification of deformation processes using machine learning in a large area of approximately 17,500 km^2 in southeastern Spain. This region encompasses a significant part of the Region of Murcia, as well as the provinces of Alicante and Almería (Figure 1). The study area exhibits a wide range of geological materials, including predominantly metamorphic hard rocks (HR) and unconsolidated sedimentary deposits (USD) (Figure 2).

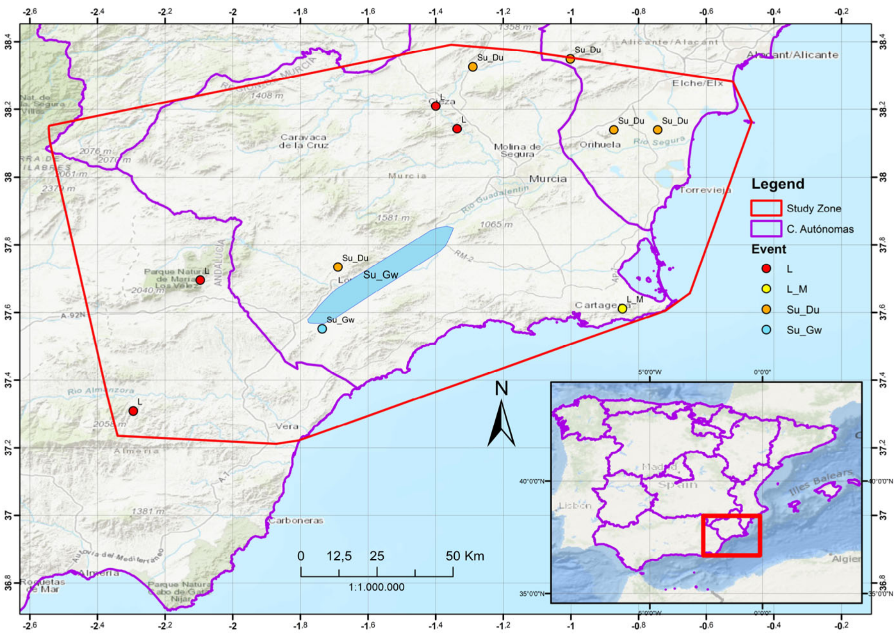

Figure 1. Localization of the study area. Four events are identifying in this zone: L_M: mining landslide; L: landslide; Su_Du: dump subsidence; Su_Gw: groundwater subsidence.

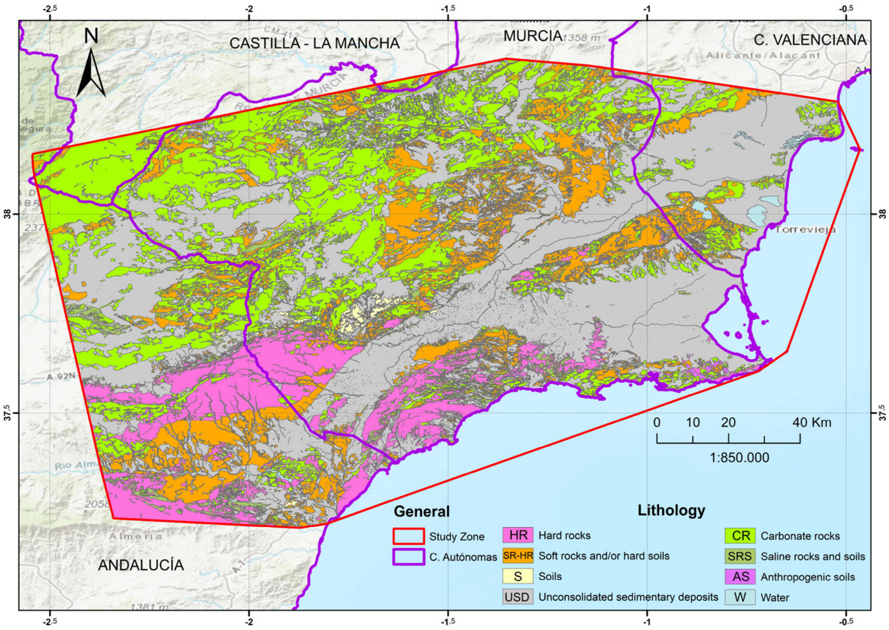

Figure 2. Geological context of the study area.

2. Methodology

Figure 3 provides an overview of the methodology employed in this research.

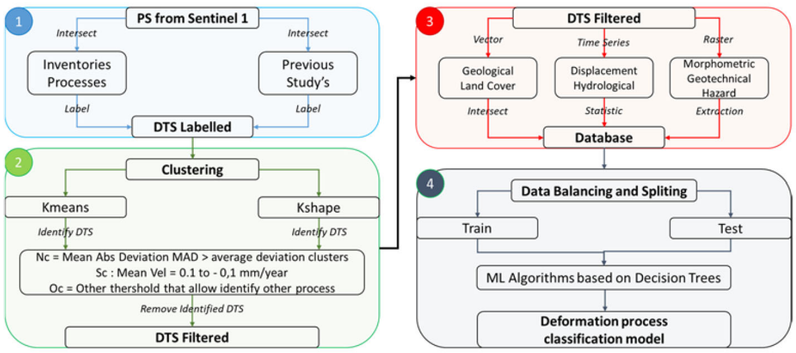

Figure 3. Methodology flowchart. (**1**) Extraction and labeling of DTS. (**2**) Clustering and filtering of DTS. (**3**) Creation of database. (**4**) Generation of classification model of deformation process using ML. Nc: Noise cluster; Sc: Stable cluster; Oc: Other clusters not related to the main process.

We utilized ground deformation measurements obtained from the processing of descending Sentinel-1 SAR data for the Region of Murcia and its surroundings, covering the period from 2015 to 2021. The selection and labeling process of each measurement point or persistent scatterer (PS hereafter), which corresponds DTS related to deformation processes, involved intersecting the DTS with national process inventories/catalogs and polygons resulting from previous SAR-based analysis and interpretation. For each labeled PS, we applied the elbow method to determine the optimal number of clusters (k) for both K-means and K-shape algorithms. The Soft_DTW algorithm served as the distance metric in both cluster analyses. We then identified and eliminated noisy and stable clusters that were not associated with deformation processes by using thresholds.

Subsequently, we constructed a database by associating the values of each of the 26 variables (Table 1) with their respective DTS. We combined the temporal information from the displacement and hydrological time series into a single aggregated variable using statistical techniques. The thematic and continuous maps were included in the database as either categorical or numerical variables. To address the issue of numerous lithological classes and prevent redundancy, we reclassified the GEODE into eight classes (Figure 2) based on their geotechnical characteristics. Finally, we trained ML algorithms based on decision trees to generate an optimal model capable of classifying DTS according to their deformation process.

Table 1. Covariates of the proposed national database, classified according to their research domain.

Displacement	Mean [1], Range [2], Desv [3], KVH [4]	Copernicus e IGME	Vector	C
Geological	Lithology [5], Age [6], Fault [7]	IGME	Vector [5,6]	N C [7]
Morphometric	Slope [8], Aspect [9], TPI [10], TWI [11], Curvature [12]	IGN	25 m [8,9,10,11]	C N [12]
Hydrological	Water Mass [13]	MITECO [13]	Vector [13]	N [13]
Geotechnical	Clay % [14], Sand % [15], Bulk Density [16], VS30 [17]	OPENGEOHUB [14,15,16], USGS [17]	30 m [14,15,16], 820 m [17]	C [14,15,16], D [17]
Hazard	Subsidence [18], Landslide [19]	IGME [18], ELSUS [19]	1 km [18], 200 m [19]	O
Land Cover/ Land Use	CLC [20], Dump [21], Quarry [22], Mining [23], Build [24], Road [25], Vol Build [26]	OPENGEOHUB [20,21,22,23], Catastro [24], IGN [25], GSHL [26]	30 m [20,21,22,23], Vector [24,25] 100 m [25,26]	N [24], C [20,21,22,23,25,26]

Each covariate is associated with a superscript number that serves to indicate features such as resolution and variable type in the table. N: nominal categorical; O: ordinal categorical; C: continue numerical; D: discrete numerical.

3. Results

During the conducted analysis, we identified a total of 58 deformation processes, with 39 corresponding to mining slides (L_M), 12 to landslides (L), 5 to dump subsidence (Su_Du), and 2 to groundwater subsidence (Su_Dw) (Figure 1a). By intersecting the data from the descending PS with the deformation processes, we successfully extracted and labeled 20,499 DTS. The vast majority of these series (97%) corresponded to subsidence caused by groundwater extraction (Su_Dw). We carried out the identification of noisy and stable time series for each deformation process through clustering of the time series. Figure 4a displays the clustering results obtained for Su_Dw. By utilizing the elbow technique, we identified six clusters. Applying thresholds related to the mean absolute deviation and mean velocity, we determined that cluster ID3 was the only one related to the deformation process. Therefore, we eliminated 5456 time series from the other clusters located at the valley edges (Figure 4b).

After filtering, we obtained 15,043 DTS related to deformation processes, which formed the database. We applied the synthetic minority over-sampling technique (SMOTE) to generate samples from minority classes and balance the data, as the majority of them belonged to the Su_DW class. We used the random forest algorithm for classification. The model achieved a perfect classification with an AUC of 1.0 in the test set, as observed in the confusion matrix of Figure 5a. Hydrological, geological, morphometric, and geotechnical variables proved to be the most relevant for the classification model (Figure 5b). Specifically, the presence of groundwater masses, distance to faults, slope, percentage of sand and clay, lithology, soil bulk density, Vs30, and geological age were the most determining variables, while variables related to displacement, hazards, and land cover had less importance.

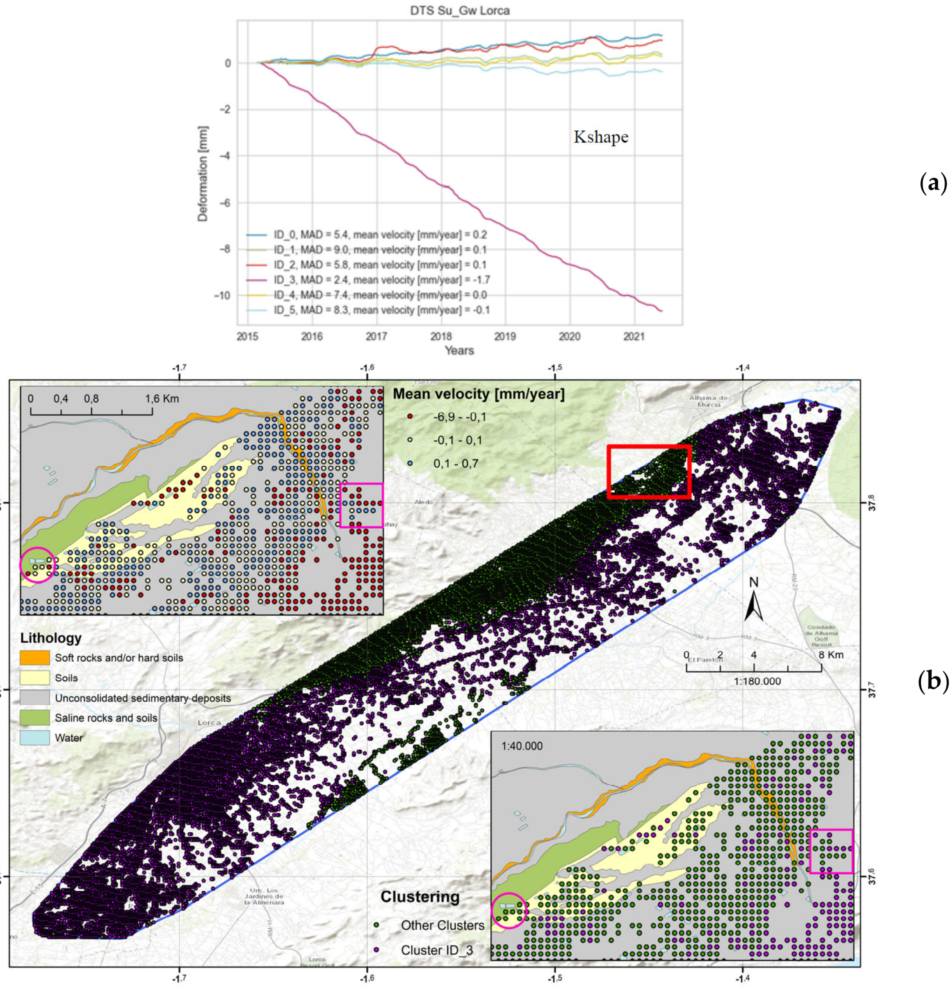

Figure 4. Results of DTS filtering. Groundwater subsidence in Lorca, SE Spain. (**a**) Centroids of clusters generated with Kshape algorithm and statistic associated with the threshold of filtering (mean absolute deviation and mean velocity). ID 3 is the unique cluster that exceeds the filtered threshold. (**b**) Spatial representation of the clustering. The comparison between the pink geometries of the corners maps allows us to identify the DTS to be removed (green points corresponding to clusters other than ID 3): noisy DTS in red (circle), stable DTS in yellow (rectangle), and DTS with inverse trends in blue (rectangle).

Figure 5. Test results of classification model. (**a**) Confusion matrix. L_M: mining landslide; L: landslide; Su_Du: dump subsidence; Su_Gw: groundwater subsidence. (**b**) Feature importance. DB = Bulk Density.

4. Discussion and Conclusions

The methodology based on statistical thresholds of DTS clusters has demonstrated its capability to identify and eliminate stable and/or noisy DTS within the same ADA. Additionally, the utilization of random forest algorithms has yielded excellent results in the classification of deformation processes when trained with displacement variables and environmental variables from the filtered DTS. Furthermore, the analysis of the database reveals that environmental variables, with the exception of land cover and hazard, exert the greatest influence on the classification of deformation processes. These variables, ranked in descending order of importance, include the presence of groundwater masses, distance to faults, slope, and percentage of sand at a 30 cm depth and lithology.

While we have obtained promising results, it is important to acknowledge that there may be an overestimation of the significance of determining variables in the classification, such as distance to faults, due to the limited spatial variability of the training data. Moreover, it is crucial to consider that the proposed methodology may yield inadequate results in identifying and classifying other types of deformation processes, such as uplift, diapirism, seismic, volcanic, and/or karstic processes. Considering these limitations, our future research will focus on further exploring the filtering methodology, variable selection, and machine learning algorithms to enable the automatic classification of various deformation processes on a national scale. We will achieve this objective by leveraging open data sources such as the European Ground Motion Service (EGMS).

Author Contributions: Conceptualization, J.R.-R., M.B.-P. and H.A.; methodology, J.R.-R., M.B.-P., H.A. and A.B.; software, J.R.-R., H.A., C.H. and A.B.; validation, M.B.-P., C.G.-A. and P.E.; formal analysis, M.B.-P., C.G.-A., P.E. and J.L.-V.; investigation, J.R.-R., A.B., R.M.M., M.C.-G., R.S., and M.M.-C.; resources, M.B.-P., P.E., C.H. and R.S.; writing—original draft preparation, J.R.-R., M.B.-P., H.A. and C.G.-A.; writing—review and editing, J.R.-R., A.B., R.M.M., M.C.-G., O.M., M.M.-C., M.C. and J.L.-V.; visualization, J.R.-R.; supervision, M.B.-P. and H.A.; project administration, C.G.-A., O.M. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been developed thanks to the pre-doctoral grant for the Training of Research Personnel (PRE2021-100044) funded by MCIN/AEI/10.13039/501100011033. It is also supported by "FSE invests in your future" within the framework of the SARAI project "Towards a smart exploitation of land displacement data for the prevention and mitigation of geological-geotechnical risks" PID2020-116540RB-C22, which is funded by MCIN/AEI/10.13039/501100011033.

Data Availability Statement: This work was based on a public dataset provided by the the Copernicus program of the European Space Agency (ESA) available at https://dataspace.copernicus.eu/ (accessed on 27 December 2023) and https://search.asf.alaska.edu/#/ (accessed on 27 December 2023). The Spanish Geological Survey (IGME) available at https://info.igme.es/cartografiadigital/ and https://info.igme.es/qafi/ (accessed on 27 December 2023). The National Geographic Institute (IGN) available at https://centrodedescargas.cnig.es (accessed on 27 December 2023). The MITECO available at https://www.miteco.gob.es/es/cartografia-y-sig.html (accessed on 27 December 2023). OPENGEOHUB available at https://opengeohub.org/ (accessed on 27 December 2023). The United State Geological Survey (USGS) available at https://earthquake.usgs.gov/data/vs30/ (accessed on 27 December 2023). The Joint Research Center (JRC) available at https://esdac.jrc.ec.europa.eu/ (accessed on 27 December 2023) and https://ghsl.jrc.ec.europa.eu/ (accessed on 27 December 2023). Cadastre available at https://www.sedecatastro.gob.es/ (accessed on 27 December 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bonì, R.; Pilla, G.; Meisina, C. Methodology for Detection and Interpretation of Ground Motion Areas with the A-DInSAR Time Series Analysis. *Remote Sens.* **2016**, *8*, 686. [CrossRef]
2. Barra, A.; Solari, L.; Béjar-Pizarro, M.; Monserrat, O.; Bianchini, S.; Herrera, G.; Crosetto, M.; Sarro, R.; González-Alonso, E.; Mateos, R.M.; et al. A Methodology to Detect and Update Active Deformation Areas Based on Sentinel-1 SAR Images. *Remote Sens.* **2017**, *9*, 1002. [CrossRef]

3. Bonì, R.; Meisina, C.; Poggio, L.; Fontana, A.; Tessari, G.; Riccardi, P.; Floris, M. Ground motion areas detection (GMA-D): An innovative approach to identify ground deformation areas using the SAR-based displacement time series. *Proc. IAHS* **2020**, *382*, 277–284. [CrossRef]
4. Navarro, J.A.; Cuevas, M.; Barra, A.; Crosetto, M. Detection of Active Deformation Areas based on Sentinel-1 imagery: An efficient, fast and flexible implementation. In Proceedings of the 18th International Scientific and Technical Conference, Crete, Greece, 24–27 September 2018.
5. Tomás, R.; Pagán, J.I.; Navarro, J.A.; Cano, M.; Pastor, J.L.; Riquelme, A.; Cuevas-González, M.; Crosetto, M.; Barra, A.; Monserrat, O.; et al. Semi-Automatic Identification and Pre-Screening of Geological–Geotechnical Deformational Processes Using Persistent Scatterer Interferometry Datasets. *Remote Sens.* **2019**, *11*, 1675. [CrossRef]
6. Festa, D.; Casagli, N.; Casu, F.; Confuorto, P.; De Luca, C.; Del Soldato, M.; Lanari, R.; Manunta, M.; Manzo, M.; Raspini, F. Automated classification of A-DInSAR-based ground deformation by using random forest. *GIScience Remote Sens.* **2022**, *59*, 1749–1766. [CrossRef]